Finding k-best MAP Solutions Using LP Relaxations

Amir Globerson
School of Computer Science and Engineering
The Hebrew University

Joint Work with: Menachem Fromer (Hebrew Univ.)

Prediction Problems

- Consider the following problem:
  - Observe variables: $\boldsymbol{x}^v$
  - Predict variables: $\boldsymbol{x}^h$

Prediction Problems

- Consider the following problem:
  - Observe variables: x^v
  - Predict variables: x^h
- Countless applications:
  - Images:
  - Error correcting codes
  - Medical diagnostics
  - Text

Visible	Hidden
Noisy Image	Source Image
Received bits	Code word
Symptoms	Disease
Sentence	Derivation

Statistical Models for Prediction

Statistical Models for Prediction

- One approach:

Statistical Models for Prediction

- One approach:
    - Assume (or learn) a model for $p(\boldsymbol{x}^h, \boldsymbol{x}^v)$

Statistical Models for Prediction

- One approach:
  - Assume (or learn) a model for $p(\boldsymbol{x}^h, \boldsymbol{x}^v)$
  - Predict the most likely hidden values

$$\arg\max_{\boldsymbol{x}^h} p(\boldsymbol{x}^h | \boldsymbol{x}^v)$$

Statistical Models for Prediction

- One approach:
  - Assume (or learn) a model for $p(\boldsymbol{x}^h, \boldsymbol{x}^v)$
  - Predict the most likely hidden values

$$\arg\max_{\boldsymbol{x}^h} p(\boldsymbol{x}^h | \boldsymbol{x}^v)$$

  - This conditional distribution often corresponds to a graphical model

Statistical Models for Prediction

- One approach:
  - Assume (or learn) a model for $p(\boldsymbol{x}^h, \boldsymbol{x}^v)$
  - Predict the most likely hidden values

$$\arg\max_{\boldsymbol{x}^h} p(\boldsymbol{x}^h | \boldsymbol{x}^v)$$

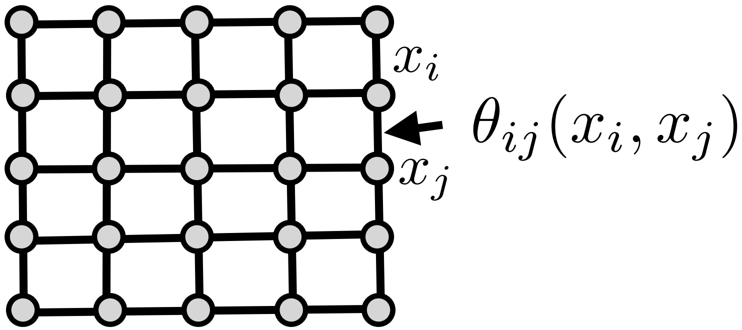
  - This conditional distribution often corresponds to a graphical model
  - Need to know how to find an assignment with maximum probability

The MAP Problem

- Given a graphical model over $x_1, \ldots, x_n$

$$p(\boldsymbol{x}) = \frac{1}{Z} e^{f(\boldsymbol{x})}$$

$$f(\boldsymbol{x}) = \sum_{ij} \theta_{ij}(x_i, x_j)$$



- Find the most likely assignment: $\arg\max_{\boldsymbol{x}} f(\boldsymbol{x})$

MAP Approximations

- $\mathbf{x}$ is discrete so generally NP hard

MAP Approximations

- $\mathbf{x}$ is discrete so generally NP hard
- Many approximation approaches:

MAP Approximations

- **x** is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search

MAP Approximations

- **x** is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search
  - Loopy belief propagation (e.g., max product)

MAP Approximations

- **x** is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search
  - Loopy belief propagation (e.g., max product)
  - Linear programming relaxations

MAP Approximations

- **x** is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search
  - Loopy belief propagation (e.g., max product)
  - Linear programming relaxations

MAP Approximations

- $\mathbf{x}$ is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search
  - Loopy belief propagation (e.g., max product)
  - Linear programming relaxations
- LP approaches

MAP Approximations

- **x** is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search
  - Loopy belief propagation (e.g., max product)
  - Linear programming relaxations
- LP approaches
  - Provide optimality certificates

MAP Approximations

- **x** is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search
  - Loopy belief propagation (e.g., max product)
  - Linear programming relaxations
- LP approaches
  - Provide optimality certificates
  - Optimal in some cases (e.g., submodular functions)

MAP Approximations

- **x** is discrete so generally NP hard
- Many approximation approaches:
  - Greedy search
  - Loopy belief propagation (e.g., max product)
  - Linear programming relaxations
- LP approaches
  - Provide optimality certificates
  - Optimal in some cases (e.g., submodular functions)
  - Can be solved via message passing

The k-best MAP Problem

The k-best MAP Problem

- Find the k best assignments for f(**x**)

The k-best MAP Problem

- Find the k best assignments for f(**x**)
- Denote these by $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}$

The k-best MAP Problem

- Find the k best assignments for f(**x**)
- Denote these by $x^{(1)}, \ldots, x^{(k)}$
- Useful in:

The k-best MAP Problem

- Find the k best assignments for f(**x**)
- Denote these by $x^{(1)}, \ldots, x^{(k)}$
- Useful in:
  - Finding multiple candidate solutions when the energy function is not accurate (e.g., protein design)

The k-best MAP Problem

- Find the k best assignments for f(**x**)
- Denote these by $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}$
- Useful in:
  - Finding multiple candidate solutions when the energy function is not accurate (e.g., protein design)
  - As a first processing stage before applying more complex methods

The k-best MAP Problem

- Find the k best assignments for f(**x**)
- Denote these by $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}$
- Useful in:
  - Finding multiple candidate solutions when the energy function is not accurate (e.g., protein design)
  - As a first processing stage before applying more complex methods
  - Supervised learning

From 2 to k best

- We can show that given a polynomial algorithm for k=2, the problem can be solved for any k in O(k)
- Focus on k=2
- Our key question: what is the LP formulation of the problem, and its relaxations?

Outline

- LP formulation of the MAP problem
- LP for 2nd best
  - General (intractable) exact formulation
  - Tractable formulation for tree graphs
  - Approximations for non-tree graphs
- Experiments

MAP and LP

MAP and LP

- MAP: $$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

MAP and LP

- MAP: $\max_{\boldsymbol{x}} f(\boldsymbol{x})$

- MAP as LP:

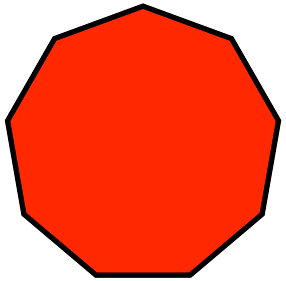
MAP and LP

- MAP: $\max_{\boldsymbol{x}} f(\boldsymbol{x})$
- MAP as LP: $\max_{\boldsymbol{\mu} \in \mathcal{S}} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$

MAP and LP

- MAP: $$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

- MAP as LP: $$\max_{\boldsymbol{\mu} \in \mathcal{S}} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$



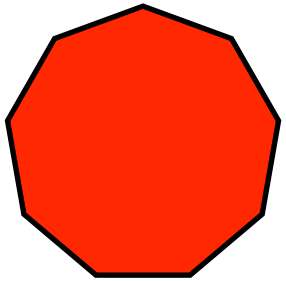
$\mathcal{S}$

MAP and LP

- MAP: $$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

- MAP as LP: $$\max_{\boldsymbol{\mu} \in \mathcal{S}} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

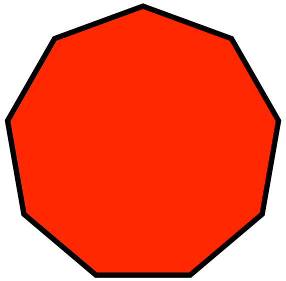
Hard



$\mathcal{S}$

MAP and LP

- MAP: $$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$
- MAP as LP: $$\max_{\boldsymbol{\mu}\in\mathcal{S}} \boldsymbol{\mu}\cdot\boldsymbol{\theta}$$ **Hard**



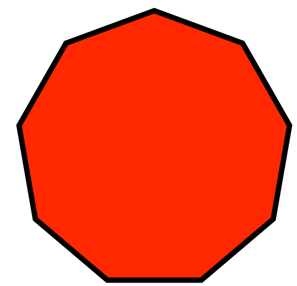
$\mathcal{S}$

- Approximate MAP via LP

MAP and LP

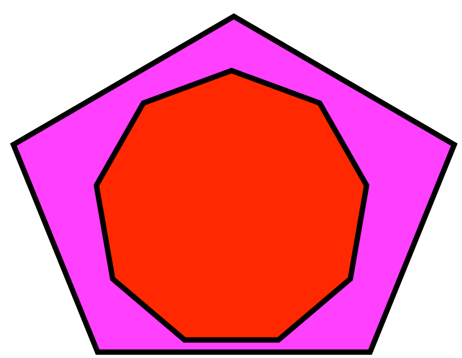
- MAP: $$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$
- MAP as LP: $$\max_{\boldsymbol{\mu}\in\mathcal{S}} \boldsymbol{\mu}\cdot\boldsymbol{\theta}$$

Hard



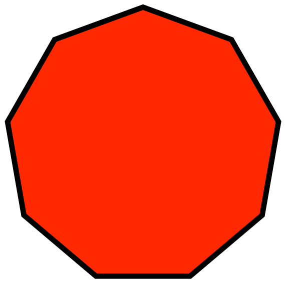
$\mathcal{S}$

- Approximate MAP via LP



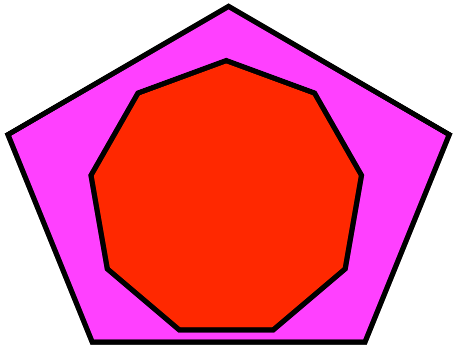
MAP and LP

- MAP: $\max_{\boldsymbol{x}} f(\boldsymbol{x})$
- MAP as LP: $\max_{\mu \in \mathcal{S}} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$ **Hard**



$\mathcal{S}$

- Approximate MAP via LP



Schlesinger, Deza & Laurent, Boros, Wainwright, Kolmogorov

LP Formulation of MAP

LP Formulation of MAP

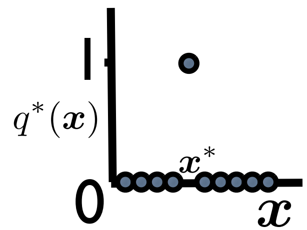
$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$

LP Formulation of MAP

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$

$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j)$$

LP Formulation of MAP

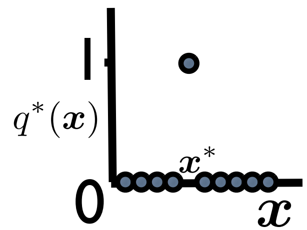


$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$

$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j)$$

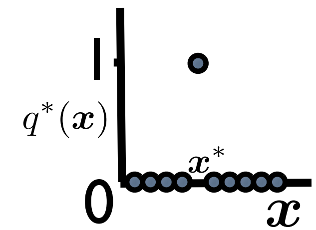
LP Formulation of MAP

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$



$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{q(\boldsymbol{x})} \sum_{ij} \sum_{x_i, x_j} q_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

LP Formulation of MAP

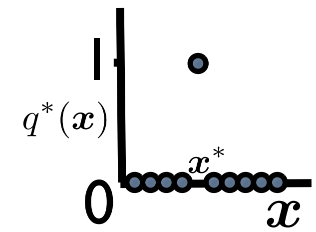


$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$

$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{q(\boldsymbol{x})} \sum_{ij} \sum_{x_i, x_j} q_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

- Objective depends only on pairwise marginals

LP Formulation of MAP

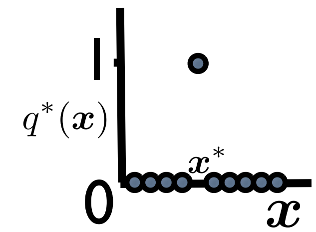


$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$

$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{q(\boldsymbol{x})} \sum_{ij} \sum_{x_i, x_j} q_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$

- Objective depends only on pairwise marginals
- But only those that correspond to some distribution $q(x)$

LP Formulation of MAP



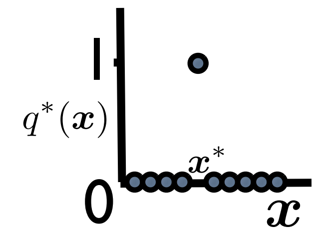
$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$

$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{q(\boldsymbol{x})} \sum_{ij} \sum_{x_i, x_j} q_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

- Objective depends only on pairwise marginals
- But only those that correspond to some distribution $q(x)$
- This set is called the Marginal polytope (Wainwright & Jordan)

LP Formulation of MAP

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$



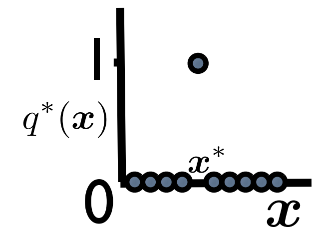
$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{q(\boldsymbol{x})} \sum_{ij} \sum_{x_i, x_j} q_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

- Objective depends only on pairwise marginals
- But only those that correspond to some distribution $q(x)$
- This set is called the Marginal polytope (Wainwright & Jordan)

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij} \mu_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

LP Formulation of MAP

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$



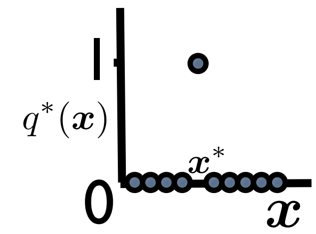
$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{q(\boldsymbol{x})} \sum_{ij} \sum_{x_i, x_j} q_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

- Objective depends only on pairwise marginals
- But only those that correspond to some distribution $q(x)$
- This set is called the Marginal polytope (Wainwright & Jordan)

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij} \mu_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

LP Formulation of MAP

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sum_{ij \in E} \theta_{ij}(x_i, x_j)$$



$$= \max_{q(\boldsymbol{x})} \sum_{\boldsymbol{x}} q(\boldsymbol{x}) \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{q(\boldsymbol{x})} \sum_{ij} \sum_{x_i, x_j} q_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

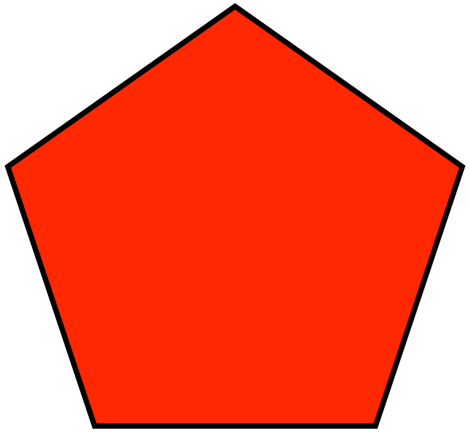
- Objective depends only on pairwise marginals
- But only those that correspond to some distribution $q(x)$
- This set is called the Marginal polytope (Wainwright & Jordan)

See: Cut polytope (Deza, Laurent),
Quadric polytope (Boros)

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij} \mu_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

The Marginal Polytope

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$

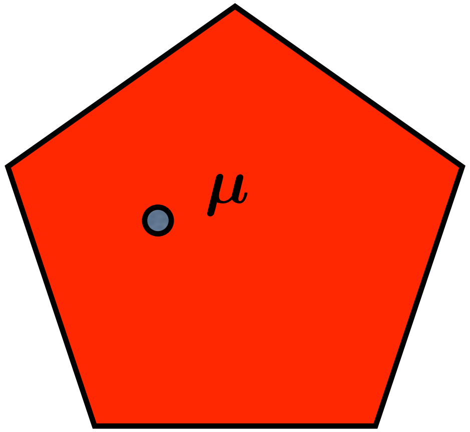


$\mathcal{M}(G)$

Marginal Polytope

The Marginal Polytope

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$

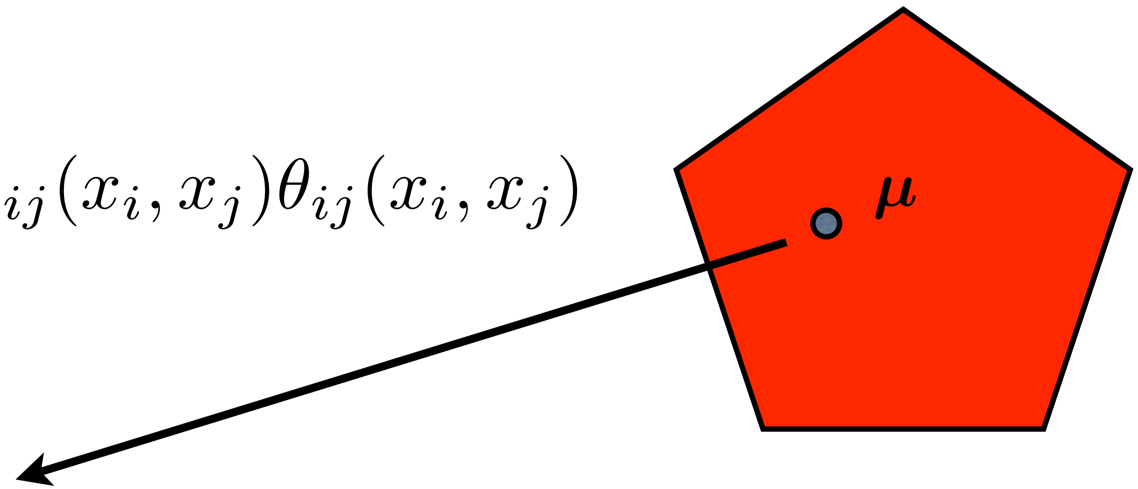


$\mathcal{M}(G)$

Marginal Polytope

The Marginal Polytope

$$\max_{\boldsymbol{\mu}\in\mathcal{M}(G)} \sum_{ij\in E}\sum_{x_i,x_j}\mu_{ij}(x_i,x_j)\theta_{ij}(x_i,x_j)$$



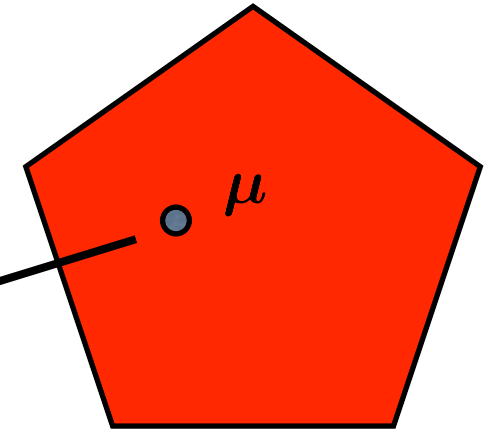
$\mathcal{M}(G)$

Marginal Polytope

- There exists a p(x) s.t. $p(x_i,x_j) = \mu_{ij}(x_i,x_j)$

The Marginal Polytope

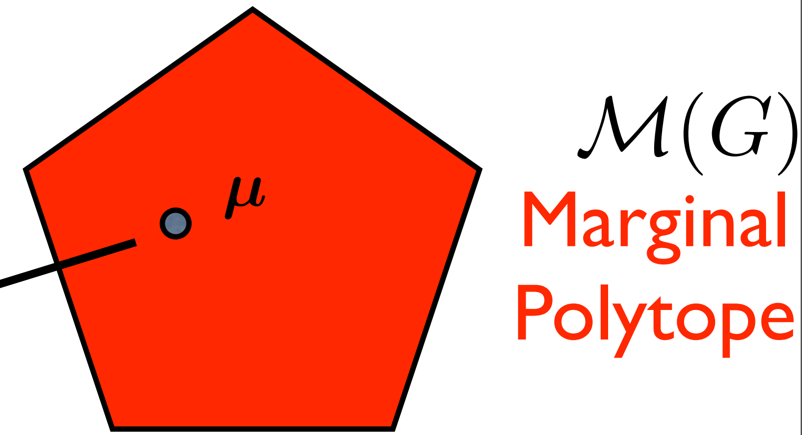
$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$



- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$
- Difficult set to characterize. Easy to outer bound

The Marginal Polytope

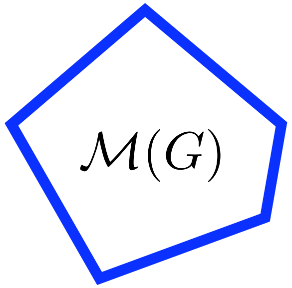
$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$



- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$
- Difficult set to characterize. Easy to outer bound
- The vertices have integral values and correspond to assignments on **x**

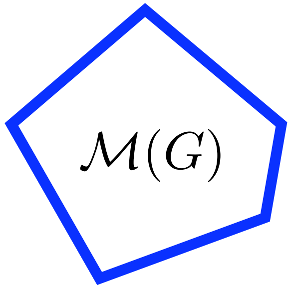
Relaxing the MAP LP

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$



Relaxing the MAP LP

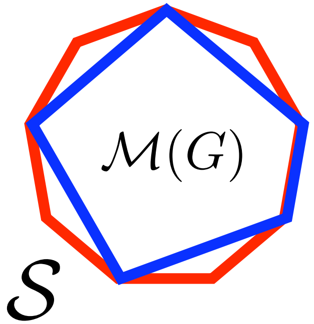
$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$



Exact but Hard!

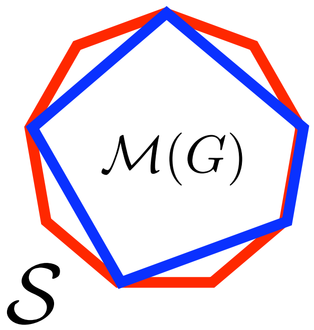
Relaxing the MAP LP

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) \leq \max_{\boldsymbol{\mu} \in \mathcal{S}} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$



Relaxing the MAP LP

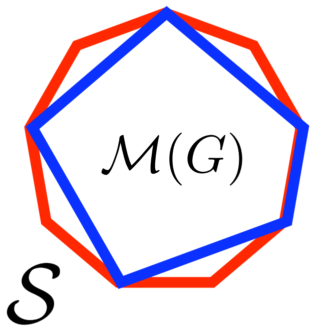
$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) \leq \max_{\boldsymbol{\mu} \in \mathcal{S}} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j)\theta_{ij}(x_i, x_j)$$



- If optimum is an integral vertex, MAP is solved

Relaxing the MAP LP

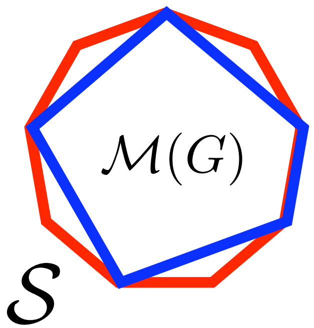
$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) \leq \max_{\boldsymbol{\mu} \in \mathcal{S}} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$



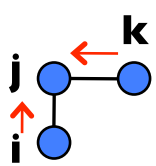
- If optimum is an integral vertex, MAP is solved
- Possible outer bound: Pairwise consistency

Relaxing the MAP LP

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) \le \max_{\boldsymbol{\mu} \in \mathcal{S}} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$



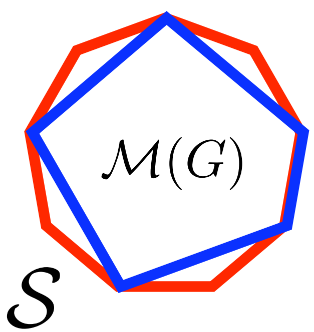
- If optimum is an integral vertex, MAP is solved
- Possible outer bound: Pairwise consistency



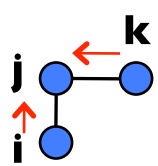
$$\sum_{x_i} \mu_{ij}(x_i, x_j) = \sum_{x_k} \mu_{jk}(x_j, x_k)$$

Relaxing the MAP LP

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) \leq \max_{\boldsymbol{\mu} \in \mathcal{S}} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$

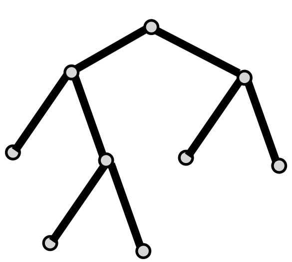


- If optimum is an integral vertex, MAP is solved
- Possible outer bound: Pairwise consistency



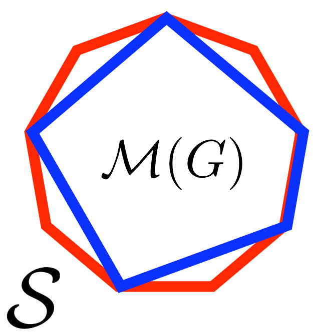
$$\sum_{x_i} \mu_{ij}(x_i, x_j) = \sum_{x_k} \mu_{jk}(x_j, x_k)$$

Exact for trees

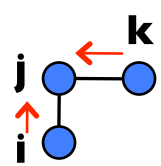


Relaxing the MAP LP

$$\max_{\boldsymbol{x}} \sum_{ij} \theta_{ij}(x_i, x_j) \leq \max_{\boldsymbol{\mu} \in \mathcal{S}} \sum_{ij \in E} \sum_{x_i, x_j} \mu_{ij}(x_i, x_j) \theta_{ij}(x_i, x_j)$$

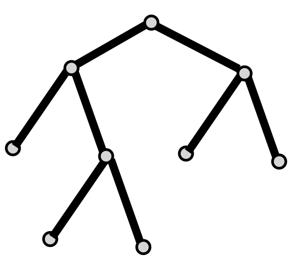


- If optimum is an integral vertex, MAP is solved
- Possible outer bound: Pairwise consistency



$$\sum_{x_i} \mu_{ij}(x_i, x_j) = \sum_{x_k} \mu_{jk}(x_j, x_k)$$

Exact for trees



- Efficient message passing schemes for solving the resulting (dual) LP

Outline

- LP formulation of the MAP problem
- LP for 2nd best
  - General (intractable) exact formulation
  - Tractable formulation for tree graphs
  - Approximations for non-tree graphs
- Experiments

The 2nd best problem and LP

MAP

2nd best

The 2nd best problem and LP

MAP

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

2nd best

The 2nd best problem and LP

MAP

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

2nd best

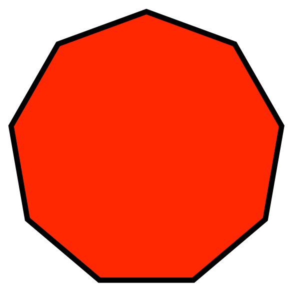
$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x})$$

The 2nd best problem and LP

<u>MAP</u>

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$



<u>2nd best</u>

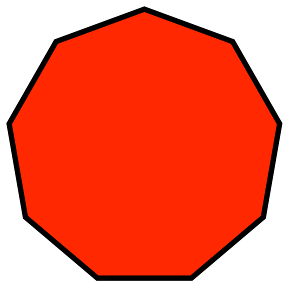
$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x})$$

The 2nd best problem and LP

MAP

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

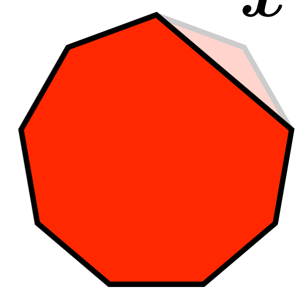


2nd best

$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G, \boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

$\boldsymbol{x}^{(1)}$

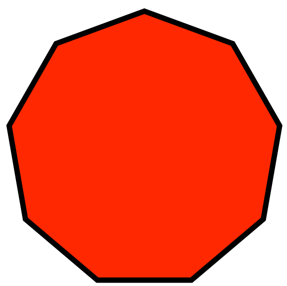


The 2nd best problem and LP

MAP

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

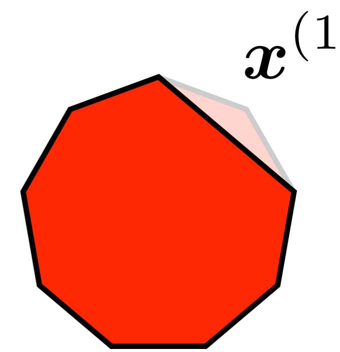


Approximations:

2nd best

$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G, \boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

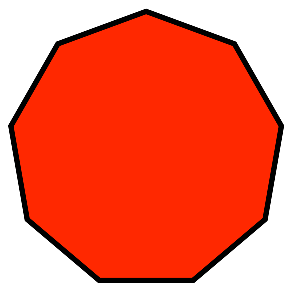


The 2nd best problem and LP

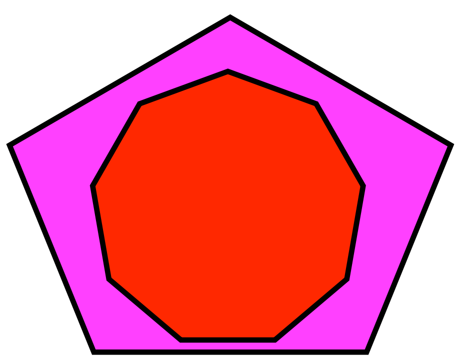
MAP

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$



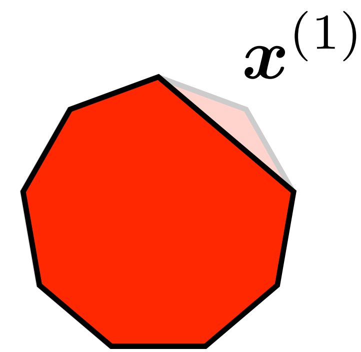
Approximations:



2nd best

$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G, \boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

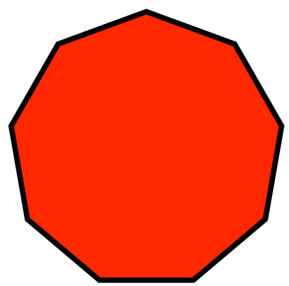


The 2$^{\text{nd}}$ best problem and LP

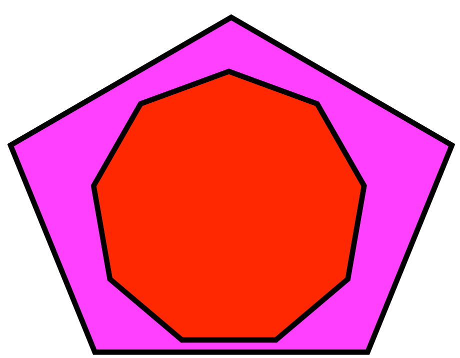
MAP

$$\max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G)} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$



Approximations:

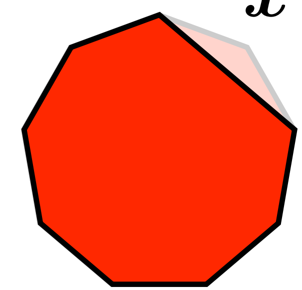


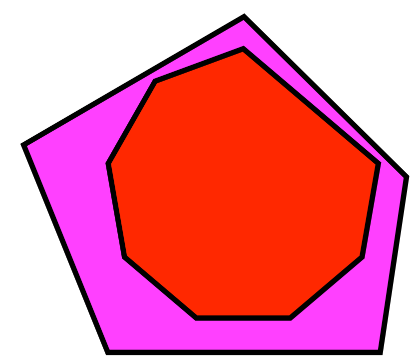
2$^{\text{nd}}$ best

$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x})$$

$$\max_{\boldsymbol{\mu} \in \mathcal{M}(G, \boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

$\boldsymbol{x}^{(1)}$



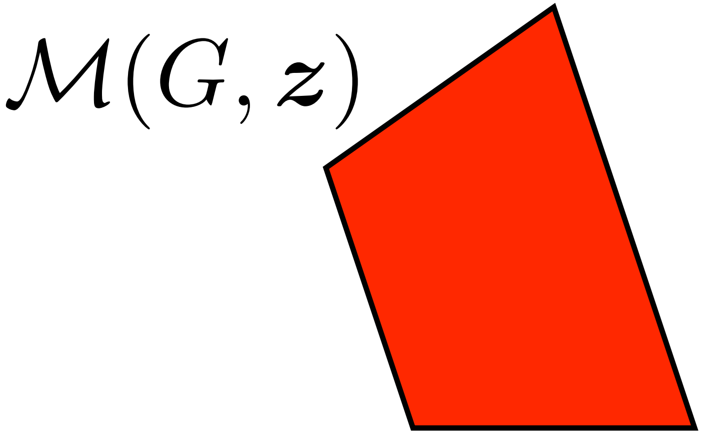


A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, z)$

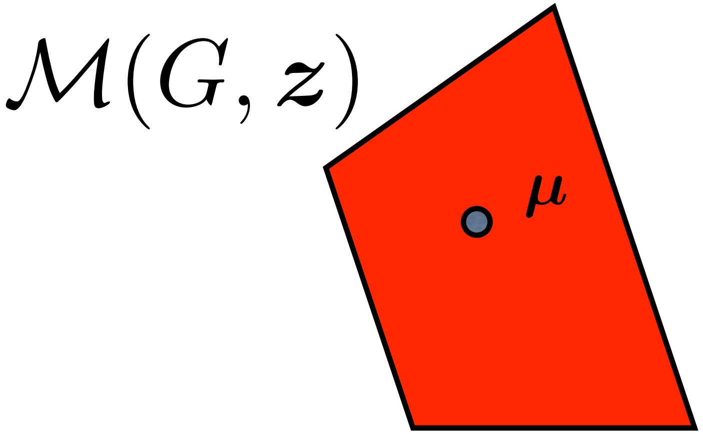
A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, z)$



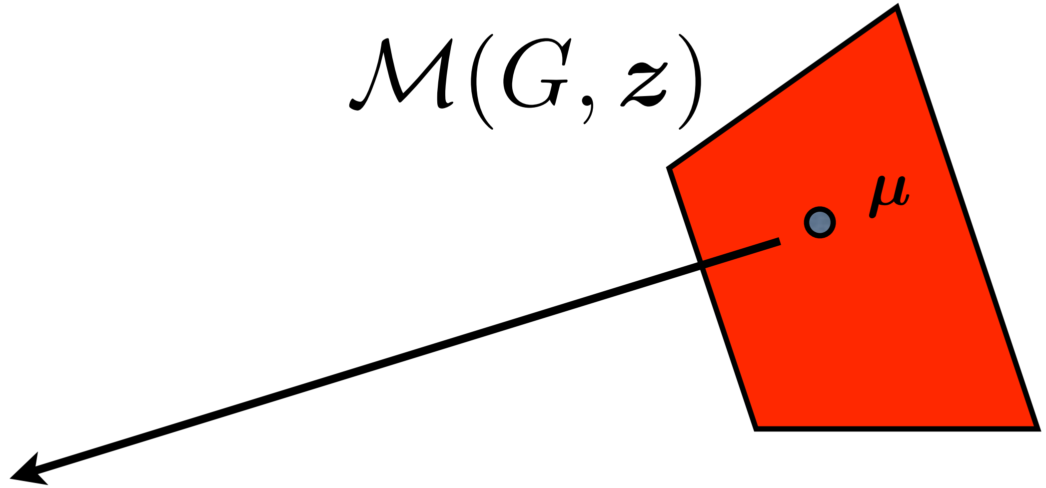
A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, z)$



A new marginal polytope

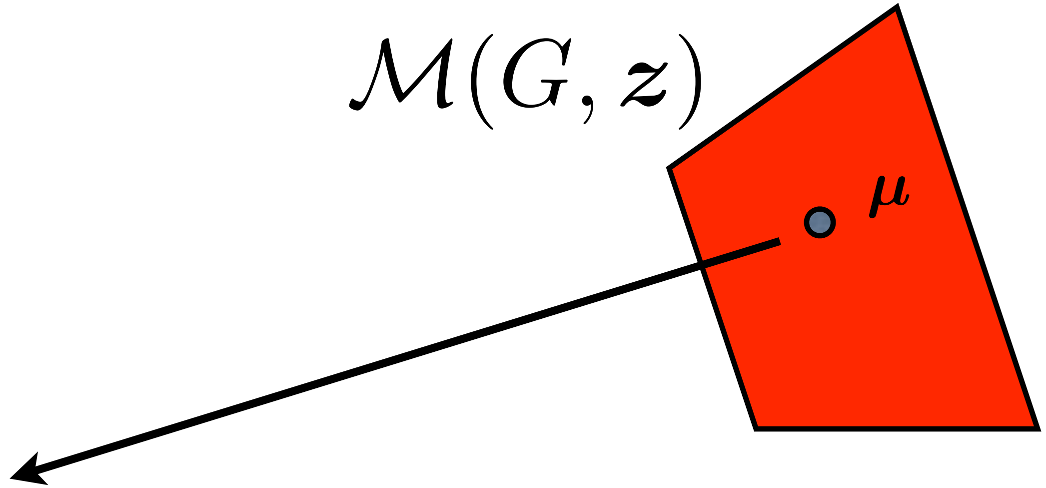
- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, \boldsymbol{z})$



- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$

A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, \boldsymbol{z})$

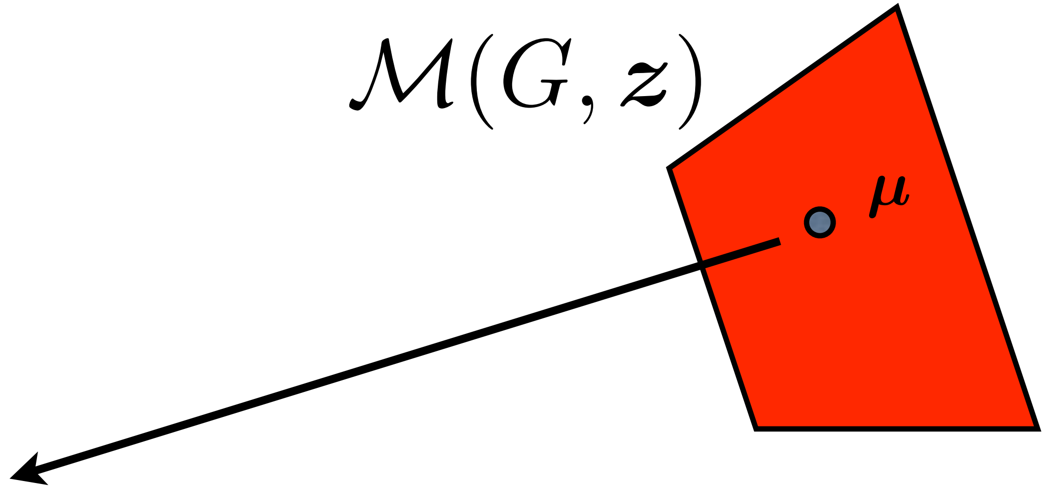


- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$

and:

A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, \boldsymbol{z})$

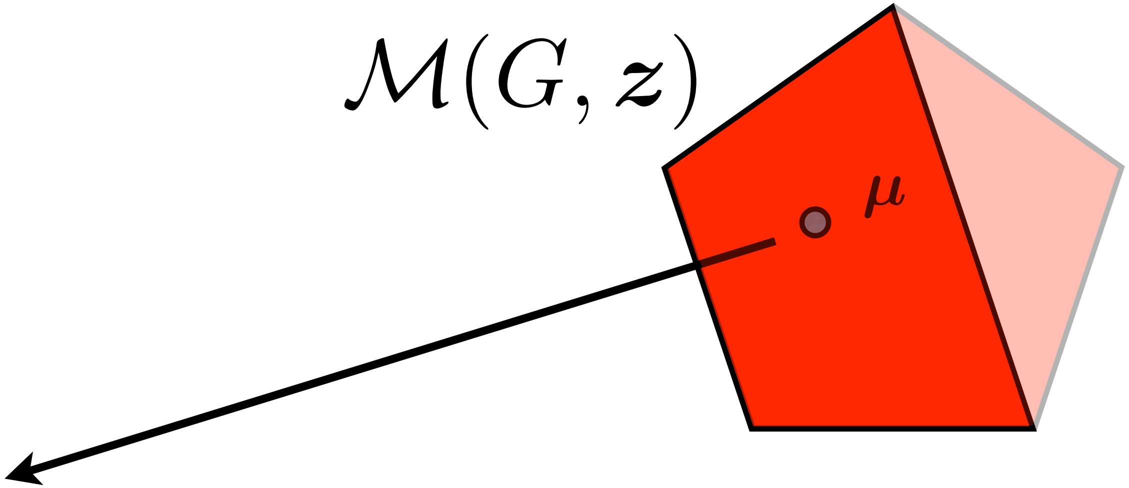


- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$

  and: $p(\boldsymbol{z}) = 0$

A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, \boldsymbol{z})$

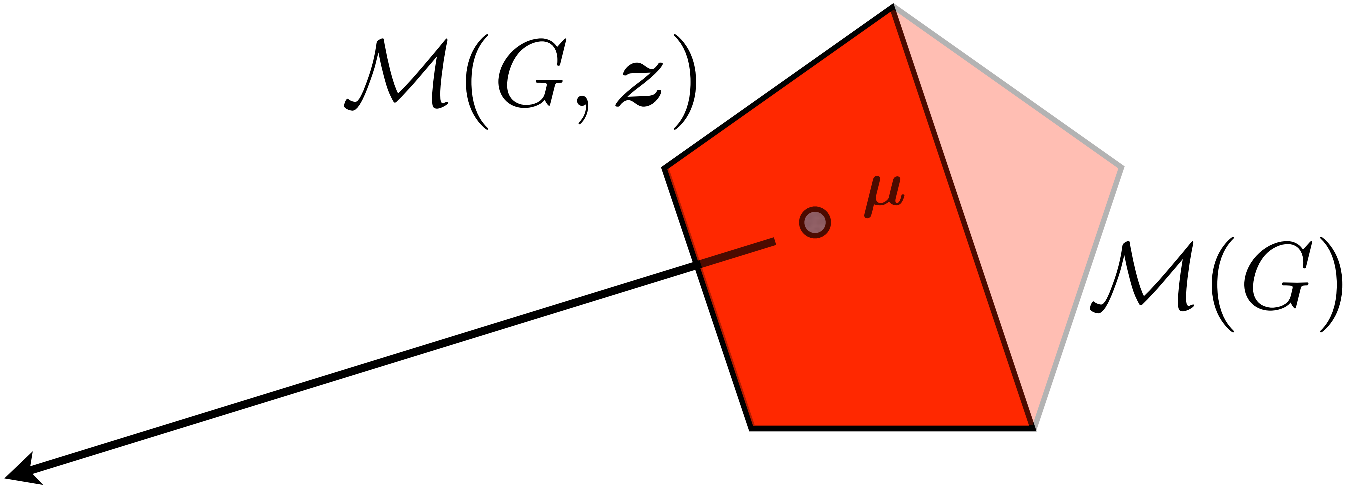


- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$

  and: $p(\boldsymbol{z}) = 0$

A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, \boldsymbol{z})$

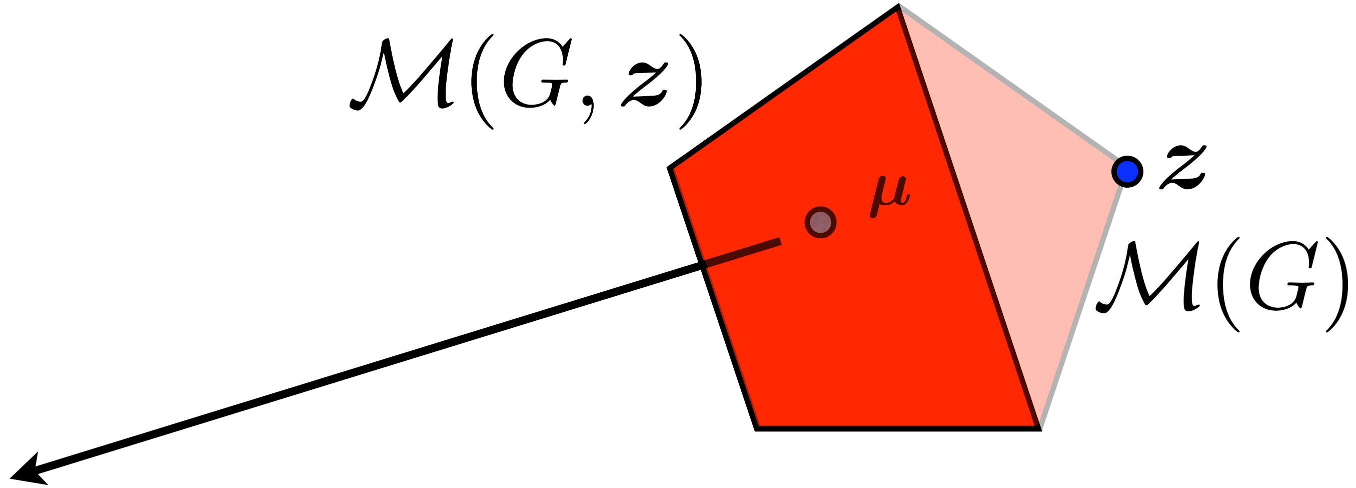


- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$

  and: $p(\boldsymbol{z}) = 0$

A new marginal polytope

- Given an assignment **z**, define the Assignment Excluding Marginal Polytope: $\mathcal{M}(G, \boldsymbol{z})$



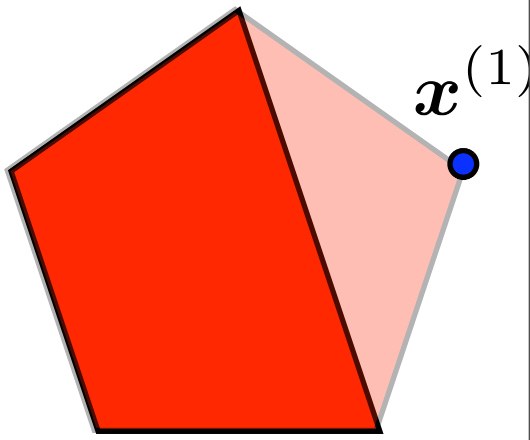
- There exists a p(x) s.t. $p(x_i, x_j) = \mu_{ij}(x_i, x_j)$

  and: $p(\boldsymbol{z}) = 0$

LP for the 2nd best problem

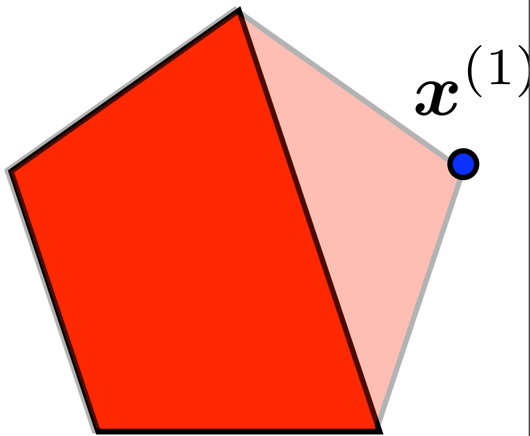
- The 2nd best problem corresponds to the following LP:

$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x}; \boldsymbol{\theta}) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G, \boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$



LP for the 2nd best problem

- The 2nd best problem corresponds to the following LP:

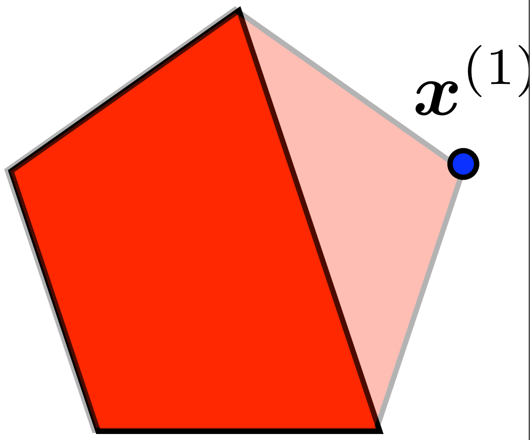


$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x}; \boldsymbol{\theta}) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G, \boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

- Is there a simple characterization of $\mathcal{M}(G, \boldsymbol{x}^{(1)})$?

LP for the 2nd best problem

- The 2nd best problem corresponds to the following LP:



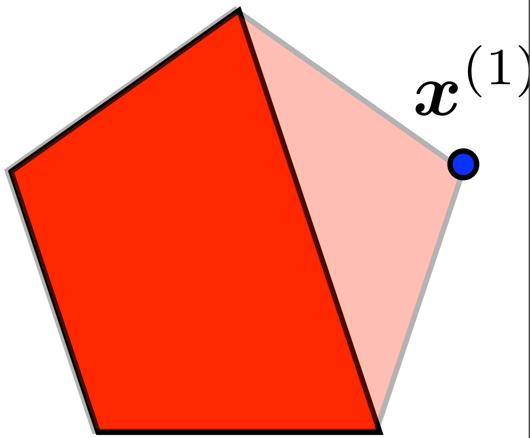
$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x};\boldsymbol{\theta}) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G,\boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$

- Is there a simple characterization of $\mathcal{M}(G,\boldsymbol{x}^{(1)})$?
- Is it $\mathcal{M}(G)$ plus one inequality?

LP for the 2nd best problem

- The 2nd best problem corresponds to the following LP:

$$\max_{\boldsymbol{x} \neq \boldsymbol{x}^{(1)}} f(\boldsymbol{x}; \boldsymbol{\theta}) = \max_{\boldsymbol{\mu} \in \mathcal{M}(G, \boldsymbol{x}^{(1)})} \boldsymbol{\mu} \cdot \boldsymbol{\theta}$$



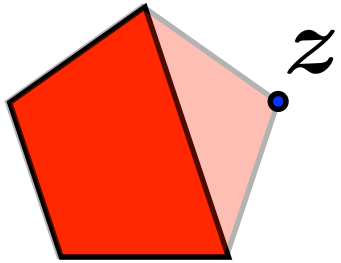
- Is there a simple characterization of $\mathcal{M}(G, \boldsymbol{x}^{(1)})$?
- Is it $\mathcal{M}(G)$ plus one inequality?
- If so, what inequality?

Outline

- LP formulation of the MAP problem
- LP for 2^{nd} best
  - General (intractable) exact formulation
  - Tractable formulation for tree graphs
  - Approximations for non-tree graphs
- Experiments

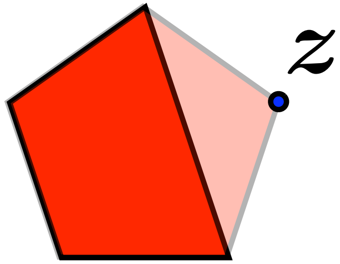
Adding inequalities to $\mathcal{M}(G)$

z

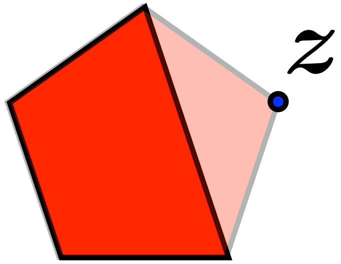


Adding inequalities to $\mathcal{M}(G)$

- Any valid inequality must separate z from the other vertices

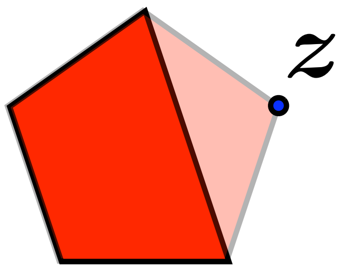


Adding inequalities to $\mathcal{M}(G)$



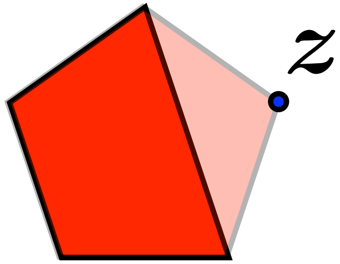
- Any valid inequality must separate z from the other vertices
- How about: $\sum_i \mu_i(z_i) \leq n-1$ (Santos 91)

Adding inequalities to $\mathcal{M}(G)$



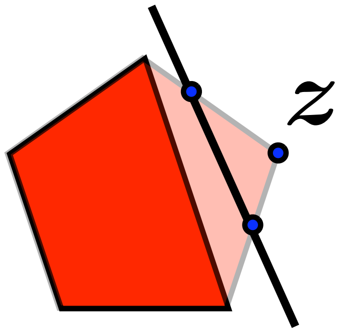
- Any valid inequality must separate z from the other vertices
- How about: $\sum_i \mu_i(z_i) \leq n-1$ (Santos 91)
- RHS is n for **z** and $n-1$ or less for other vertices

Adding inequalities to $\mathcal{M}(G)$



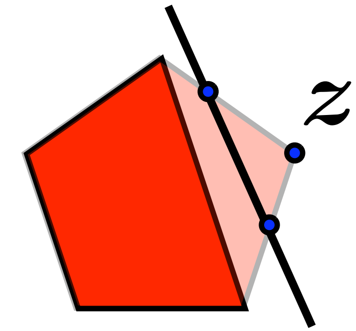
- Any valid inequality must separate z from the other vertices
- How about: $\sum_i \mu_i(z_i) \leq n-1$ (Santos 91)
- RHS is n for **z** and $n-1$ or less for other vertices
- But: Results in fractional vertices, even for trees

Adding inequalities to $\mathcal{M}(G)$



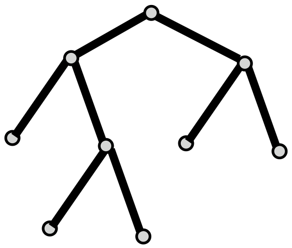
- Any valid inequality must separate z from the other vertices
- How about: $\sum_i \mu_i(z_i) \leq n-1$ (Santos 91)
- RHS is n for **z** and $n-1$ or less for other vertices
- But: Results in fractional vertices, even for trees

Adding inequalities to $\mathcal{M}(G)$



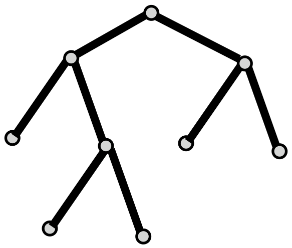
- Any valid inequality must separate z from the other vertices
- How about: $\sum_i \mu_i(z_i) \leq n - 1$ (Santos 91)
- RHS is n for **z** and $n - 1$ or less for other vertices
- But: Results in fractional vertices, even for trees
- Only an outer bound on $\mathcal{M}(G, z)$

The tree case



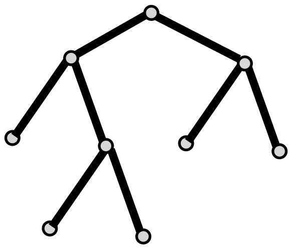
The tree case

- Focus on the case where G is a tree



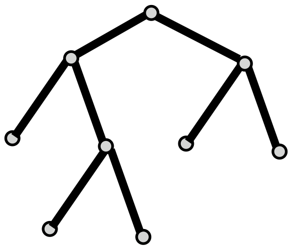
The tree case

- Focus on the case where G is a tree
- $\mathcal{M}(G)$ is given by pairwise consistency



The tree case

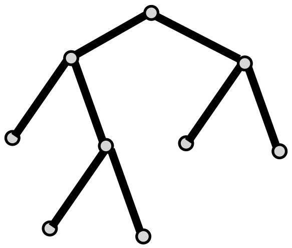
- Focus on the case where G is a tree
- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:



$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

The tree case

- Focus on the case where G is a tree
- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:

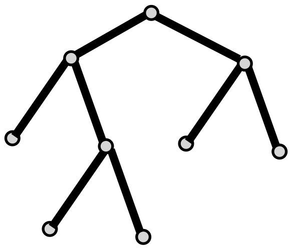


$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

Bethe: $H(\boldsymbol{\mu}) = \sum_i (1 - d_i) H_i(X_i) + \sum_{ij \in G} H(X_i, X_j)$

The tree case

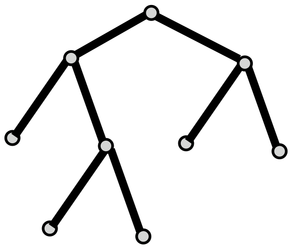
- Focus on the case where G is a tree
- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:



$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

The tree case

- Focus on the case where G is a tree
- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:



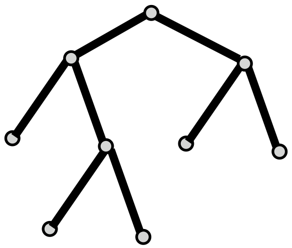
$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

- <u>Theorem:</u>

$$\mathcal{M}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$$

The tree case

- Focus on the case where G is a tree

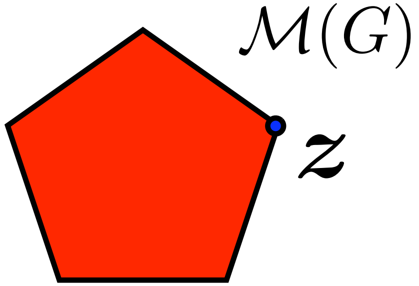


- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:

$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

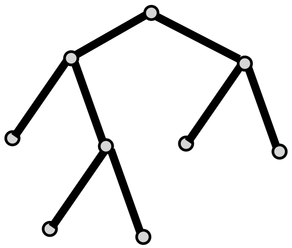
- <u>Theorem:</u>

$$\mathcal{M}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \ \ \boldsymbol{\mu} \in \mathcal{M}(G), \ \ I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$$



The tree case

- Focus on the case where G is a tree



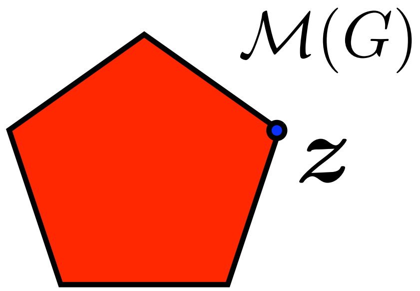
- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:

$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

$$I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$$

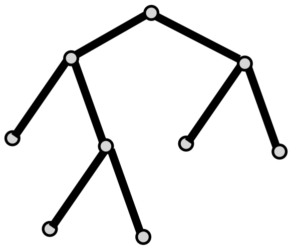
- <u>Theorem:</u>

$$\mathcal{M}(G, \boldsymbol{z}) = \{ \boldsymbol{\mu} \mid \ \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$$



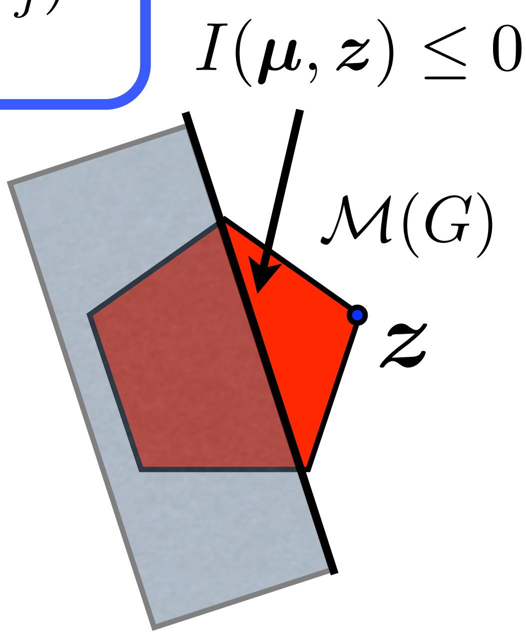
The tree case

- Focus on the case where G is a tree



- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:

$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

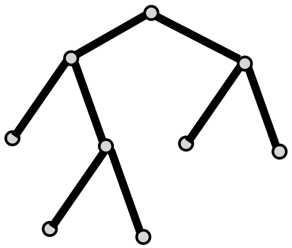


- <u>Theorem:</u>

$$\mathcal{M}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$$

The tree case

- Focus on the case where G is a tree

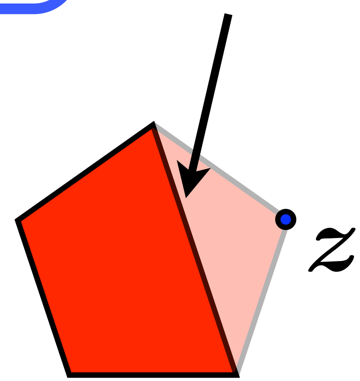


- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:

$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

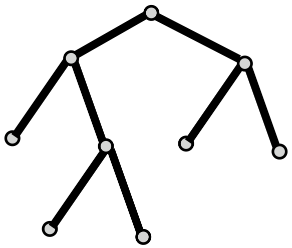
- <u>Theorem:</u>

$$\mathcal{M}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \ \ \boldsymbol{\mu} \in \mathcal{M}(G), \ \ \ I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$$



The tree case

- Focus on the case where G is a tree

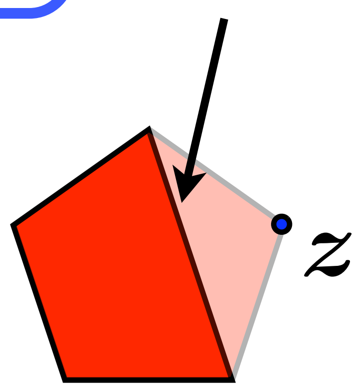


- $\mathcal{M}(G)$ is given by pairwise consistency
- Define:

$$\boxed{I(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in G} \mu_{ij}(z_i, z_j)}$$

- <u>Theorem:</u>

$$\mathcal{M}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \ \boldsymbol{\mu} \in \mathcal{M}(G), \ \ I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$$



Proof...

Proof

- Define: $\mathcal{A}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$

Proof

- Define: $\mathcal{A}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$
- Want to show: $\mathcal{A}(G, \boldsymbol{z}) = \mathcal{M}(G, \boldsymbol{z})$

Proof

- Define: $\mathcal{A}(G,\boldsymbol{z}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu},\boldsymbol{z}) \leq 0 \ \}$
- Want to show that if $\boldsymbol{\mu} \in \mathcal{A}(G,\boldsymbol{z})$ there exists a p(**x**) that has these marginals and p(**z**)=0.

Proof

- Define: $\mathcal{A}(G,z) = \{\boldsymbol{\mu} \mid \ \boldsymbol{\mu} \in \mathcal{M}(G), \ \ I(\boldsymbol{\mu}, z) \leq 0 \ \}$
- Want to show that if $\mu \in \mathcal{A}(G,z)$ there exists a p(**x**) that has these marginals and p(**z**)=0.

Can construct p(x)

Proof

- Define: $\mathcal{A}(G, z) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, z) \leq 0 \ \}$
- Want to show that if $\mu \in \mathcal{A}(G, z)$ there exists a p(**x**) that has these marginals and p(**z**)=0.

Proof

- Define: $\mathcal{A}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$
- Want to show that if $\boldsymbol{\mu} \in \mathcal{A}(G, \boldsymbol{z})$ there exists a p(**x**) that has these marginals and p(**z**)=0.

$$F(\boldsymbol{\mu}) = \left\{ \begin{array}{ll} \min & p(\boldsymbol{z}) \\ \mathrm{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \end{array} \right.$$

Proof

- Define: $\mathcal{A}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \ \ \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$
- Want to show that if $\boldsymbol{\mu} \in \mathcal{A}(G, \boldsymbol{z})$ there exists a p(**x**) that has these marginals and p(**z**)=0.

$$F(\boldsymbol{\mu}) = \left\{ \begin{array}{ll} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \end{array} \right. = 0$$

$$\forall \boldsymbol{\mu} \in \mathcal{A}(G, \boldsymbol{z})$$

Proof

- Define: $\mathcal{A}(G, \boldsymbol{z}) = \{\boldsymbol{\mu} \mid \ \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0 \ \}$
- Want to show that if $\boldsymbol{\mu} \in \mathcal{A}(G, \boldsymbol{z})$ there exists a p(**x**) that has these marginals and p(**z**)=0.

$$F(\boldsymbol{\mu}) = \left\{ \begin{array}{ll} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \end{array} \right. = 0 \qquad \forall \boldsymbol{\mu} \in \mathcal{A}(G, \boldsymbol{z})$$

- In fact we can show that for trees:

$$\boldsymbol{\mu} \in \mathcal{M}(G) \implies F(\boldsymbol{\mu}) = \max\{0, I(\boldsymbol{\mu}, \boldsymbol{z})\}$$

Proof - key ideas

$$F(\boldsymbol{\mu}) = \begin{cases} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \end{cases}$$

Proof - key ideas

$$F(\boldsymbol{\mu}) = \begin{cases} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \qquad \forall \boldsymbol{x} \neq \boldsymbol{z} \end{cases}$$

Proof - key ideas

$$F'(\boldsymbol{\mu}) = \begin{cases} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \qquad \forall \boldsymbol{x} \neq \boldsymbol{z} \end{cases}$$

Proof - key ideas

$$F'(\boldsymbol{\mu}) = \left\{ \begin{array}{ll} \min & p(\boldsymbol{z}) \\ \text{s.t.} & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \qquad \forall \boldsymbol{x} \neq \boldsymbol{z} \end{array} \right.$$

Dual:

$$\begin{array}{ll} \max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\ \text{s.t.} & \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\ & \sum_{ij} \lambda_{ij}(\boldsymbol{z}_i, \boldsymbol{z}_j) + \sum_i \lambda_i(\boldsymbol{z}_i) = 1 \end{array}$$

Proof - key ideas

$$F'(\boldsymbol{\mu}) = \begin{cases} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \qquad \forall \boldsymbol{x} \neq \boldsymbol{z} \end{cases}$$

Dual:

$$\begin{array}{ll} \max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\ \text{s.}t. & \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\ & \sum_{ij} \lambda_{ij}(\boldsymbol{z}_i, \boldsymbol{z}_j) + \sum_i \lambda_i(\boldsymbol{z}_i) = 1 \end{array}$$

- We show that the value of the above is $I(\boldsymbol{\mu}, \boldsymbol{z})$

Proof - key ideas

$$F'(\boldsymbol{\mu}) = \left\{ \begin{array}{ll} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \qquad\qquad \forall \boldsymbol{x} \neq \boldsymbol{z} \end{array} \right.$$

Dual:

$$\begin{array}{ll} \max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\ \text{s.}t. & \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\ & \sum_{ij} \lambda_{ij}(\boldsymbol{z}_i, \boldsymbol{z}_j) + \sum_i \lambda_i(\boldsymbol{z}_i) = 1 \end{array}$$

- We show that the value of the above is $I(\boldsymbol{\mu}, \boldsymbol{z})$
- From there it's easy to conclude that

Proof - key ideas

$$F'(\boldsymbol{\mu}) = \begin{cases} \min & p(\boldsymbol{z}) \\ \text{s.}t. & p_{ij}(x_i, x_j) = \mu_{ij}(x_i, x_j) \\ & p_i(x_i) = \mu_i(x_i) \\ & p(\boldsymbol{x}) \geq 0 \qquad \forall \boldsymbol{x} \neq \boldsymbol{z} \end{cases}$$

Dual:

$$\begin{array}{ll} \max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\ \text{s.}t. & \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\ & \sum_{ij} \lambda_{ij}(\boldsymbol{z}_i, \boldsymbol{z}_j) + \sum_i \lambda_i(\boldsymbol{z}_i) = 1 \end{array}$$

- We show that the value of the above is $I(\boldsymbol{\mu}, \boldsymbol{z})$
- From there it's easy to conclude that

$$F(\boldsymbol{\mu}) = \max\{0, I(\boldsymbol{\mu}, \boldsymbol{z})\}$$

Proof - Max marginals

$$
\begin{array}{ll}
\max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\
\mathrm{s}.t. & \lambda(\boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\
& \lambda(\boldsymbol{z}) = 1 \\
& \lambda(\boldsymbol{x}) = \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i)
\end{array}
$$

Proof - Max marginals

$$\begin{array}{ll} \max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\ \text{s.}t. & \lambda(\boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\ & \lambda(\boldsymbol{z}) = 1 \\ & \lambda(\boldsymbol{x}) = \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i) \end{array}$$

- Use max-marginals:

Proof - Max marginals

$$\begin{aligned} \max \quad & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\ \text{s.}t. \quad & \lambda(\boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\ & \lambda(\boldsymbol{z}) = 1 \\ & \lambda(\boldsymbol{x}) = \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i) \end{aligned}$$

- Use max-marginals:

$$\begin{aligned} \bar{\lambda}(x_i) &= \max_{\hat{x}:\hat{x}_i = x_i} \lambda(x) \\ \bar{\lambda}(x_i.x_j) &= \max_{\hat{x}:\hat{x}_i = x_i, \hat{x}_j = x_j} \lambda(x) \end{aligned}$$

Proof - Max marginals

$$
\begin{array}{ll}
\max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\
\text{s.}t. & \lambda(\boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\
& \lambda(\boldsymbol{z}) = 1 \\
& \lambda(\boldsymbol{x}) = \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i)
\end{array}
$$

- Use max-marginals:

$$
\begin{array}{rcl}
\bar{\lambda}(x_i) & = & \max_{\hat{x}:\hat{x}_i = x_i} \lambda(x) \\
\bar{\lambda}(x_i.x_j) & = & \max_{\hat{x}:\hat{x}_i = x_i, \hat{x}_j = x_j} \lambda(x)
\end{array}
\Longrightarrow
\begin{array}{rcll}
\bar{\lambda}(z_i) & = & 1 & \\
\bar{\lambda}(x_i) & \leq & 0 & x_i \neq z_i
\end{array}
$$

Proof - Max marginals

$$\begin{array}{ll} \max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\ \text{s.}t. & \lambda(\boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\ & \lambda(\boldsymbol{z}) = 1 \\ & \lambda(\boldsymbol{x}) = \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i) \end{array}$$

- Use max-marginals:

$$\begin{aligned} \bar{\lambda}(x_i) &= \max_{\hat{x}:\hat{x}_i = x_i} \lambda(x) \\ \bar{\lambda}(x_i.x_j) &= \max_{\hat{x}:\hat{x}_i = x_i, \hat{x}_j = x_j} \lambda(x) \end{aligned} \Rightarrow \begin{aligned} \bar{\lambda}(z_i) &= 1 \\ \bar{\lambda}(x_i) &\leq 0 \quad x_i \neq z_i \end{aligned}$$

- Rewrite: $\lambda(\boldsymbol{x}) = \sum_i (1 - d_i)\bar{\lambda}(x_i) + \sum_{ij \in T} \bar{\lambda}_{ij}(x_i, x_j)$

Proof - Max marginals

$$
\begin{array}{ll}
\max & \boldsymbol{\lambda} \cdot \boldsymbol{\mu} \\
\text{s.}t. & \lambda(\boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{x} \neq \boldsymbol{z} \\
 & \lambda(\boldsymbol{z}) = 1 \\
 & \lambda(\boldsymbol{x}) = \sum_{ij} \lambda_{ij}(x_i, x_j) + \sum_i \lambda_i(x_i)
\end{array}
$$

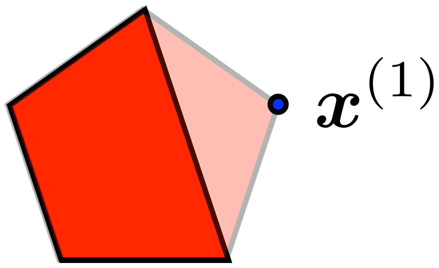
- Use max-marginals:

$$
\begin{array}{rcl}
\bar{\lambda}(x_i) & = & \max_{\hat{x}:\hat{x}_i = x_i} \lambda(x) \\
\bar{\lambda}(x_i.x_j) & = & \max_{\hat{x}:\hat{x}_i = x_i, \hat{x}_j = x_j} \lambda(x)
\end{array}
\quad \Longrightarrow \quad
\begin{array}{rcll}
\bar{\lambda}(z_i) & = & 1 & \\
\bar{\lambda}(x_i) & \leq & 0 & x_i \neq z_i
\end{array}
$$

- Rewrite: $\lambda(\boldsymbol{x}) = \sum_i (1 - d_i)\bar{\lambda}(x_i) + \sum_{ij \in T} \bar{\lambda}_{ij}(x_i, x_j)$
- Result follows after some algebra

Tree Graph - Summary

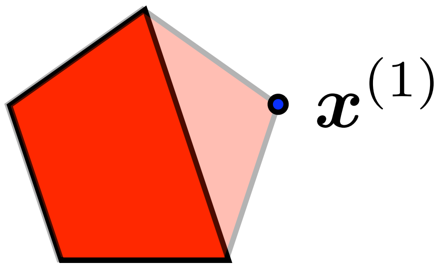
$$\mathcal{M}(G, \boldsymbol{x}^{(1)}) = \{ \boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) \leq 0 \ \}$$



Tree Graph - Summary

$$\mathcal{M}(G, \boldsymbol{x}^{(1)}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) \leq 0 \ \}$$

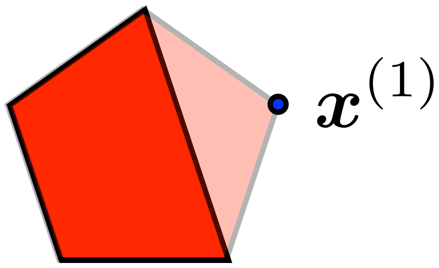
- The LP for 2nd best differs from the marginal polytope by one linear inequality constraint



Tree Graph - Summary

$$\mathcal{M}(G, \boldsymbol{x}^{(1)}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) \leq 0 \ \}$$

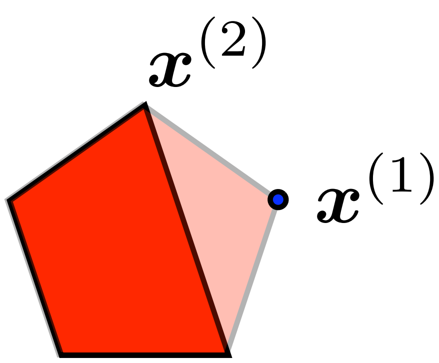
- The LP for 2nd best differs from the marginal polytope by one linear inequality constraint
- The 2nd best satisfies $I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) = 0$ so it cannot be any assignment



Tree Graph - Summary

$$\mathcal{M}(G, \boldsymbol{x}^{(1)}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) \leq 0 \ \}$$

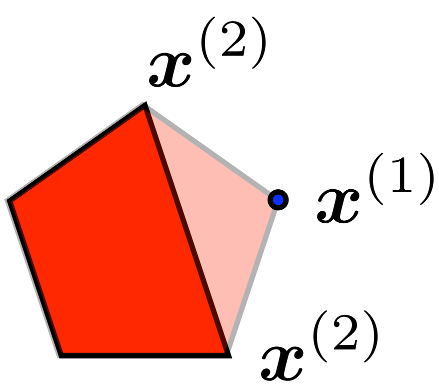
- The LP for 2nd best differs from the marginal polytope by one linear inequality constraint
- The 2nd best satisfies $I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) = 0$ so it cannot be any assignment



Tree Graph - Summary

$$\mathcal{M}(G, \boldsymbol{x}^{(1)}) = \{\boldsymbol{\mu} \mid \quad \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) \leq 0 \ \}$$

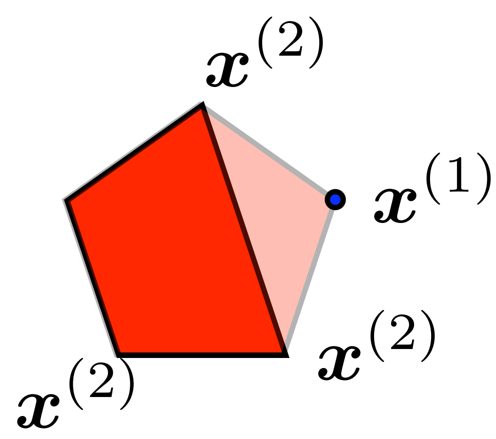
- The LP for 2nd best differs from the marginal polytope by one linear inequality constraint
- The 2nd best satisfies $I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) = 0$ so it cannot be any assignment



Tree Graph - Summary

$$\mathcal{M}(G, \boldsymbol{x}^{(1)}) = \{\boldsymbol{\mu} \mid \ \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) \leq 0 \ \}$$

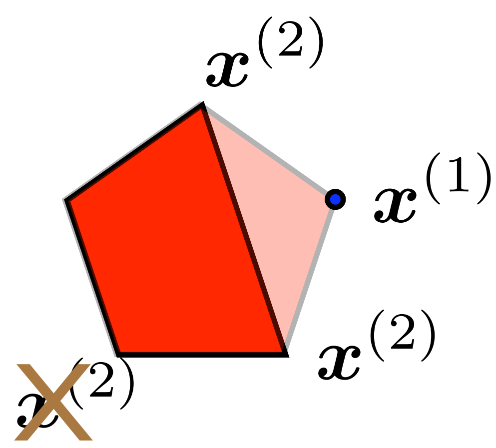
- The LP for 2nd best differs from the marginal polytope by one linear inequality constraint
- The 2nd best satisfies $I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) = 0$ so it cannot be any assignment



Tree Graph - Summary

$$\mathcal{M}(G,\boldsymbol{x}^{(1)}) = \{\boldsymbol{\mu} \mid \ \ \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu},\boldsymbol{x}^{(1)}) \le 0 \ \}$$

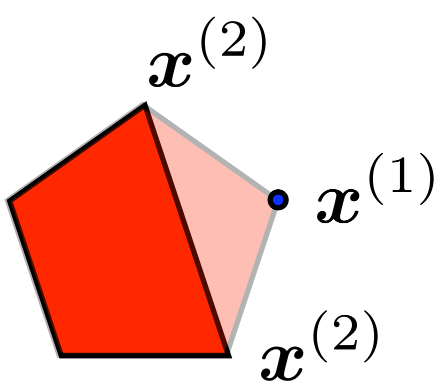
- The LP for 2nd best differs from the marginal polytope by one linear inequality constraint
- The 2nd best satisfies $I(\boldsymbol{\mu},\boldsymbol{x}^{(1)}) = 0$ so it cannot be any assignment



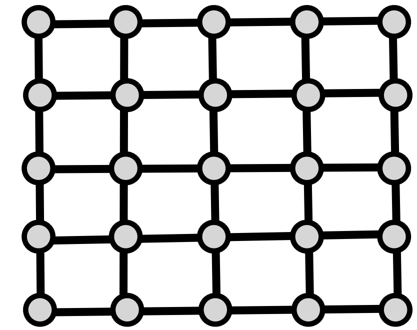
Tree Graph - Summary

$$\mathcal{M}(G, \boldsymbol{x}^{(1)}) = \{\boldsymbol{\mu} \mid \ \ \boldsymbol{\mu} \in \mathcal{M}(G), \quad I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) \leq 0 \ \}$$

- The LP for 2nd best differs from the marginal polytope by one linear inequality constraint
- The 2nd best satisfies $I(\boldsymbol{\mu}, \boldsymbol{x}^{(1)}) = 0$ so it cannot be any assignment



Non tree graphs



- Any graph can be converted into a junction tree
- We can apply our tree result there
- For a junction tree with cliques C and separators S, the inequality is:

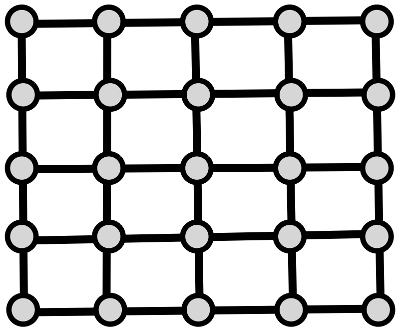
$$\sum_{S \in \mathcal{S}} (1 - d_S)\mu_S(z_S) + \sum_{C \in \mathcal{C}} \mu_C(z_C) \leq 0$$

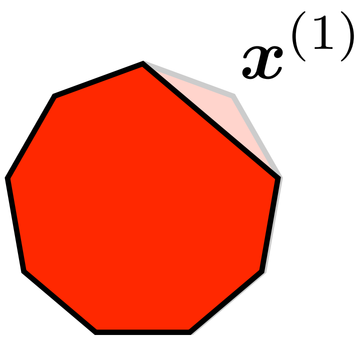
- Specifying the marginal polytope requires a number of variables exponential in the tree width. Not practical.

Outline

- LP formulation of the MAP problem
- LP for 2^{nd} best
  - General (intractable) exact formulation
  - Tractable formulation for tree graphs
  - Approximations for non-tree graphs
- Experiments

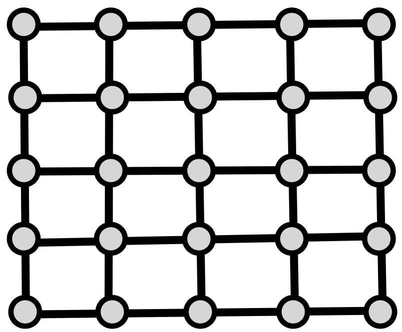
Non trees - Approximations

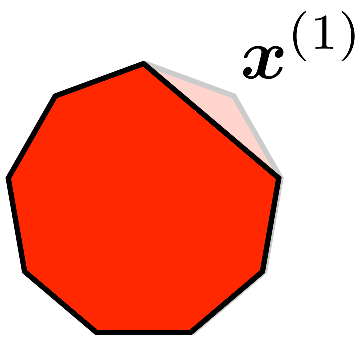


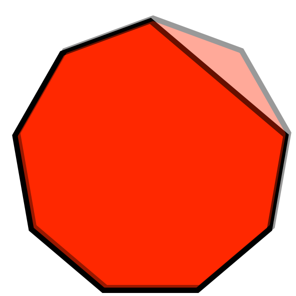


True $\mathcal{M}(G, \boldsymbol{x}^{(1)})$

Non trees - Approximations

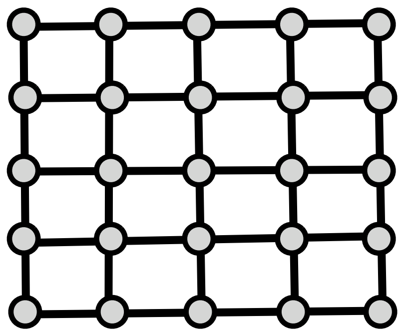




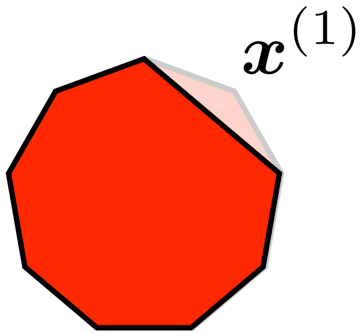


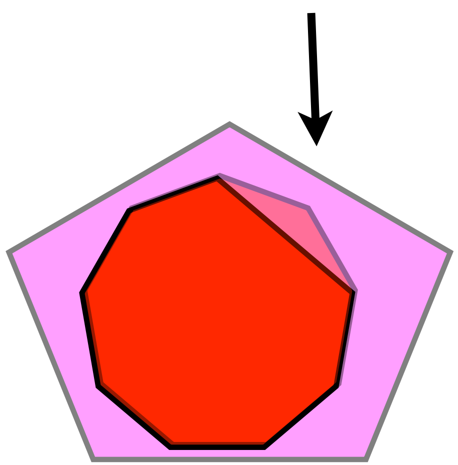
True $\mathcal{M}(G, \boldsymbol{x}^{(1)})$

Non trees - Approximations



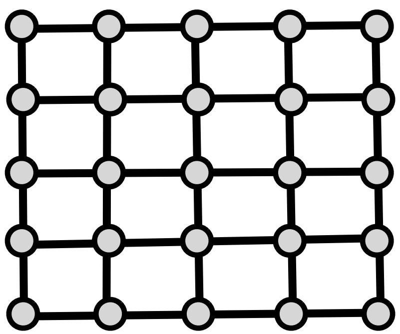
Outer bound on $\mathcal{M}(G)$



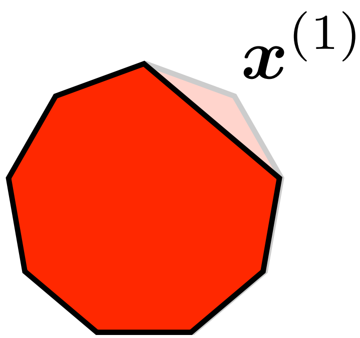


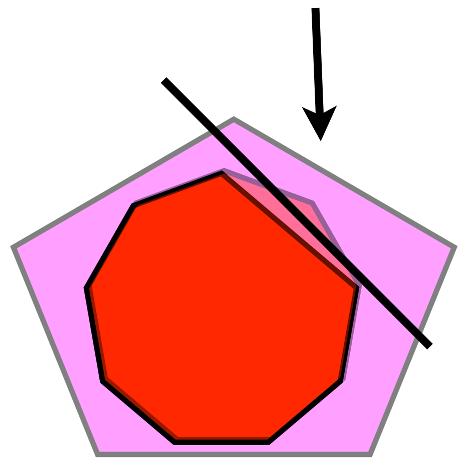
True $\mathcal{M}(G, \boldsymbol{x}^{(1)})$

Non trees - Approximations



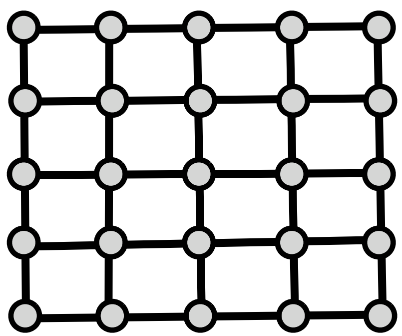
Outer bound on $\mathcal{M}(G)$



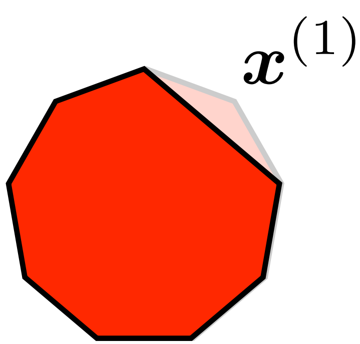


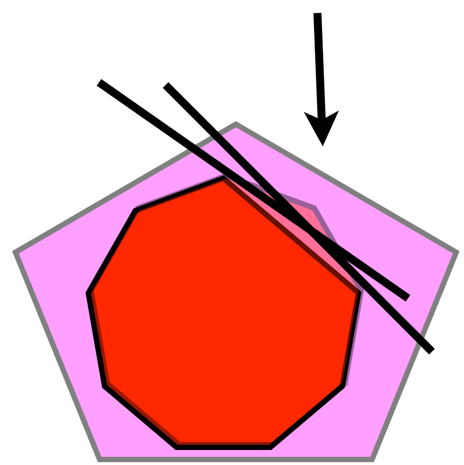
True $\mathcal{M}(G, \boldsymbol{x}^{(1)})$

Non trees - Approximations



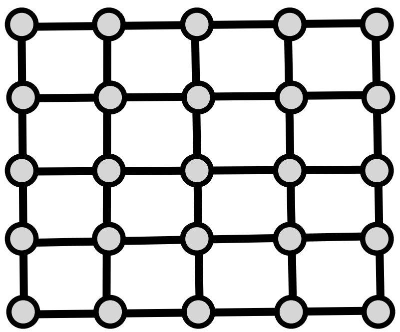
Outer bound on $\mathcal{M}(G)$



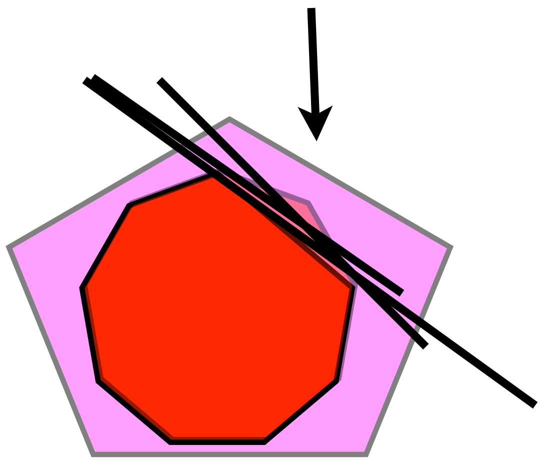


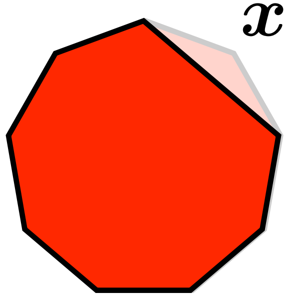
True $\mathcal{M}(G, \boldsymbol{x}^{(1)})$

Non trees - Approximations



Outer bound on $\mathcal{M}(G)$





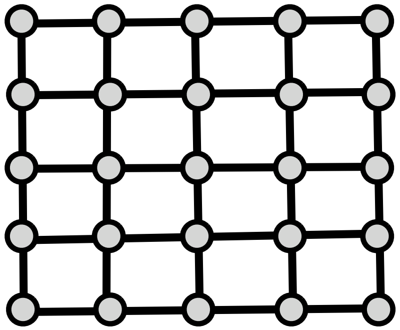
True $\mathcal{M}(G, \boldsymbol{x}^{(1)})$

Spanning tree inequalities

- Give a spanning subtree T of G define

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$

- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$

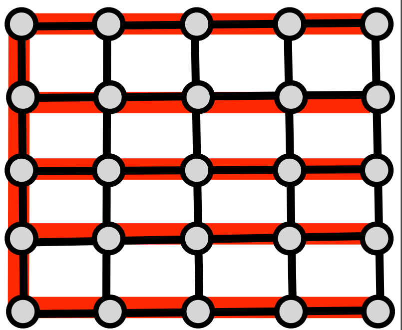


Spanning tree inequalities

- Give a spanning subtree T of G define

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$

- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$

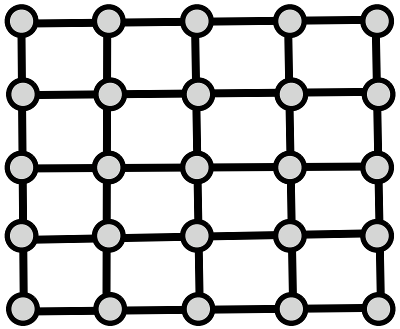


Spanning tree inequalities

- Give a spanning subtree T of G define

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$

- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$

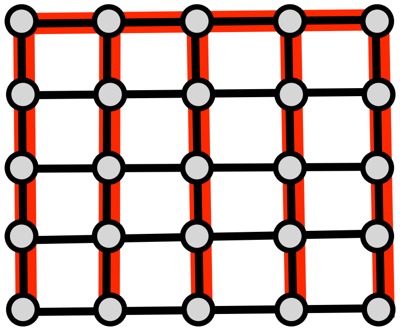


Spanning tree inequalities

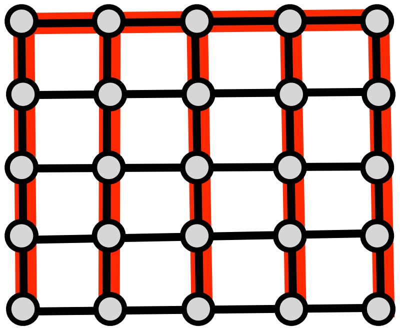
- Give a spanning subtree T of G define

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$

- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$



Spanning tree inequalities



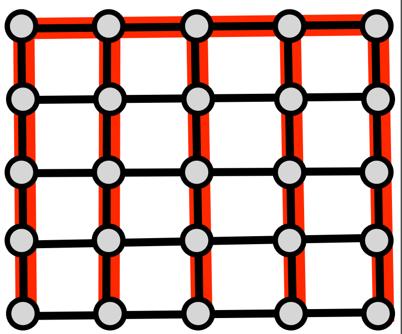
- Give a spanning subtree T of G define

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$

- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$
- Separates **z** from the other vertices but might result in fractional vertices

Spanning tree inequalities

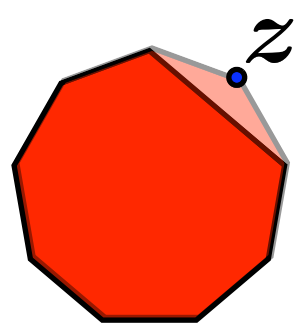
- Give a spanning subtree T of G define



$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$

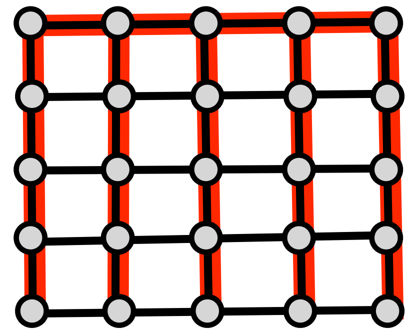
- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$

- Separates **z** from the other vertices but might result in fractional vertices



Spanning tree inequalities

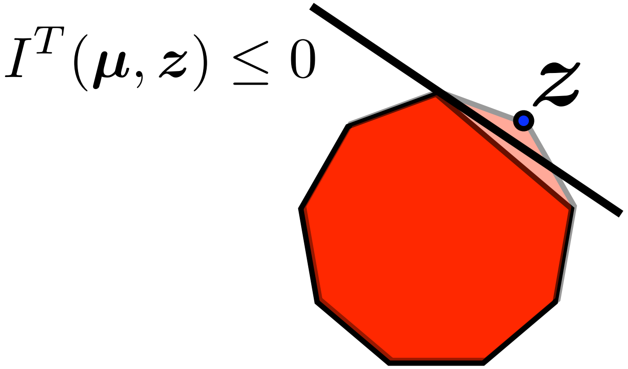
- Give a spanning subtree T of G define



$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$

- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$

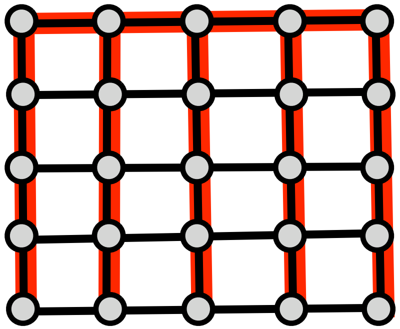
- Separates **z** from the other vertices but might result in fractional vertices



Spanning tree inequalities

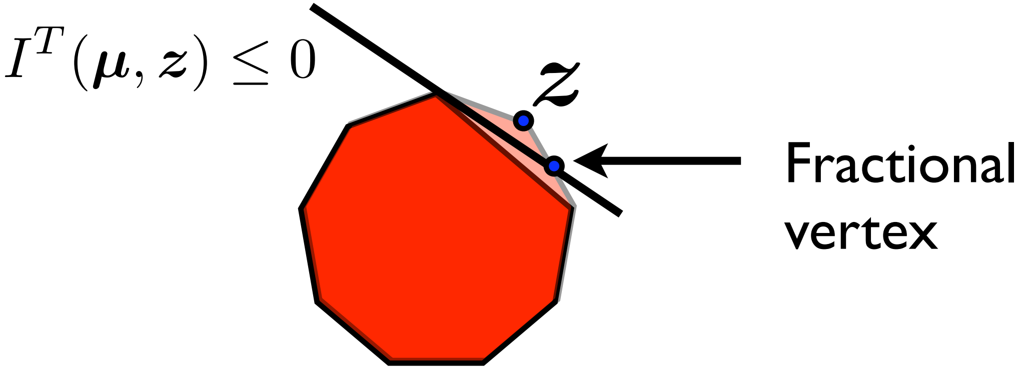
- Give a spanning subtree T of G define

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_i (1 - d_i)\mu_i(z_i) + \sum_{ij \in T} \mu_{ij}(z_i, z_j)$$



- And the constraint: $I^T(\boldsymbol{\mu}, \boldsymbol{z}) \leq 0$

- Separates **z** from the other vertices but might result in fractional vertices



Adding all spanning trees

Adding all spanning trees

- Can we add **all** spanning tree inequalities efficiently?

Adding all spanning trees

- Can we add **all** spanning tree inequalities efficiently?
- Yes, via a cutting plane approach:

Adding all spanning trees

- Can we add **all** spanning tree inequalities efficiently?
- Yes, via a cutting plane approach:
  - Start with one inequality

Adding all spanning trees

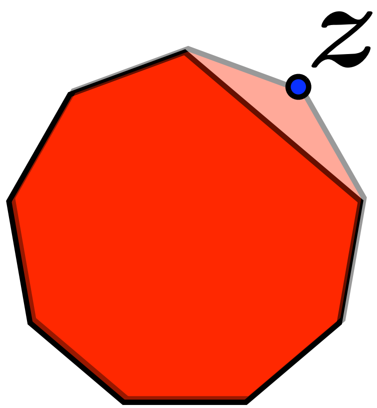
- Can we add **all** spanning tree inequalities efficiently?
- Yes, via a cutting plane approach:
  - Start with one inequality
  - Solve LP

Adding all spanning trees

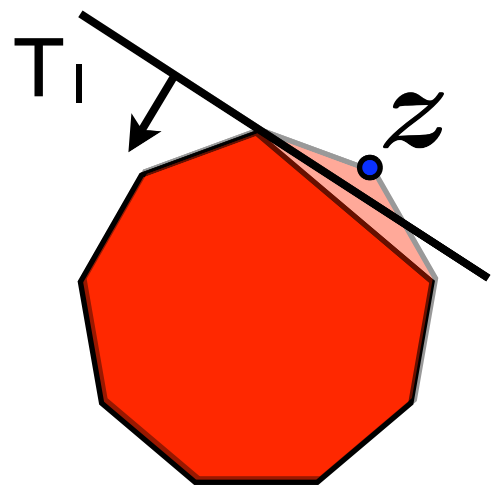
- Can we add **all** spanning tree inequalities efficiently?
- Yes, via a cutting plane approach:
  - Start with one inequality
  - Solve LP
  - If solution is fractional, find a violated tree inequality (if exists) and add it

Cutting Plane Algorithm

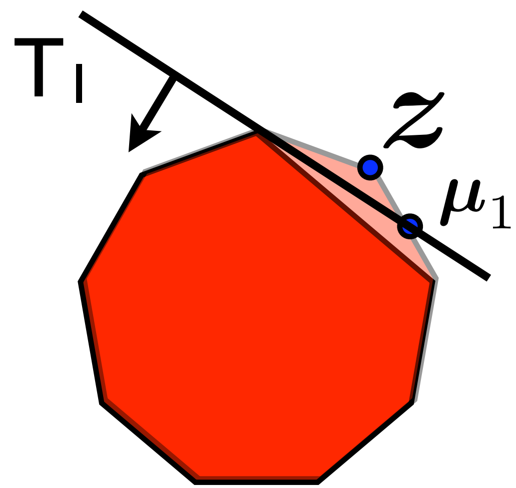
Cutting Plane Algorithm



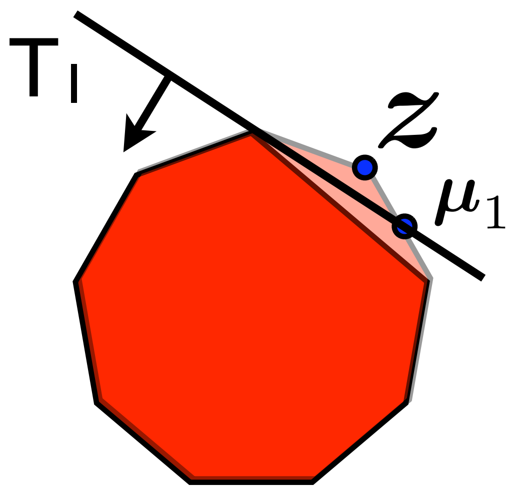
Cutting Plane Algorithm



Cutting Plane Algorithm

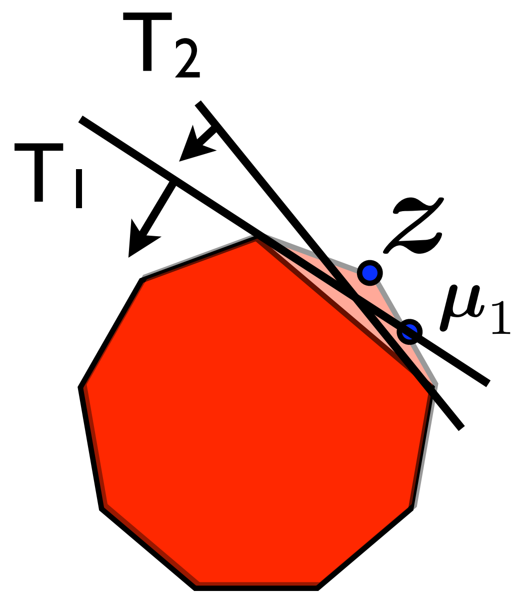


Cutting Plane Algorithm



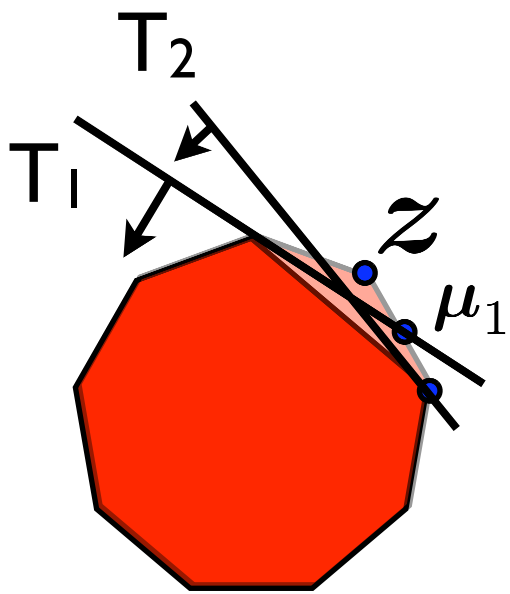
Is there a tree inequality that μ_1 violates?

Cutting Plane Algorithm



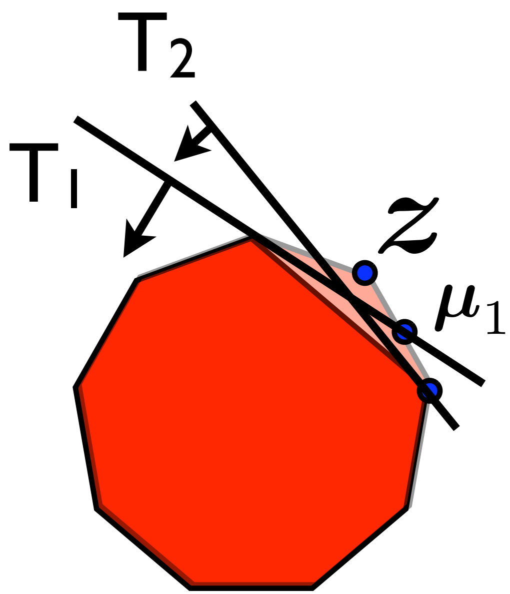
Is there a tree inequality that μ_1 violates?

Cutting Plane Algorithm



Is there a tree inequality that μ_1 violates?

Cutting Plane Algorithm



Is there a tree inequality that μ_1 violates?

- How do we find a violated tree inequality?
- Note: Even all spanning tree inequalities might not suffice

Finding a violated spanning tree

- For a given $\boldsymbol{\mu}$ find $\max_T I^T(\boldsymbol{\mu}, \boldsymbol{z})$
- If it's positive, add the maximizing tree

Finding a violated spanning tree

- For a given $\boldsymbol{\mu}$ find $\max_T I^T(\boldsymbol{\mu}, \boldsymbol{z})$
- If it's positive, add the maximizing tree
- How can we maximize over all trees? Note that:

Finding a violated spanning tree

- For a given $\boldsymbol{\mu}$ find $\max_T I^T(\boldsymbol{\mu}, \boldsymbol{z})$
- If it's positive, add the maximizing tree
- How can we maximize over all trees? Note that:

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_{ij \in T} \Big[\mu_{ij}(z_i, z_j) - \mu_i(z_i) - \mu_j(z_j)\Big] + \sum_i \mu_i(z_i)$$

Finding a violated spanning tree

- For a given $\boldsymbol{\mu}$ find $\max_T I^T(\boldsymbol{\mu}, \boldsymbol{z})$
- If it's positive, add the maximizing tree
- How can we maximize over all trees? Note that:

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_{ij \in T} \Big[\mu_{ij}(z_i, z_j) - \mu_i(z_i) - \mu_j(z_j) \Big] + \sum_i \mu_i(z_i)$$

Finding a violated spanning tree

- For a given $\boldsymbol{\mu}$ find $\max_T I^T(\boldsymbol{\mu}, \boldsymbol{z})$
- If it's positive, add the maximizing tree
- How can we maximize over all trees? Note that:

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_{ij \in T} \Big[\underbrace{\mu_{ij}(z_i, z_j) - \mu_i(z_i) - \mu_j(z_j)}_{w_{ij}} \Big] + \sum_i \mu_i(z_i)$$

Finding a violated spanning tree

- For a given $\boldsymbol{\mu}$ find $\max_T I^T(\boldsymbol{\mu}, \boldsymbol{z})$
- If it's positive, add the maximizing tree

- How can we maximize over all trees? Note that:

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_{ij \in T} \Big[\underbrace{\mu_{ij}(z_i, z_j) - \mu_i(z_i) - \mu_j(z_j)}_{w_{ij}} \Big] + \underbrace{\sum_i \mu_i(z_i)}_{\text{Fixed}}$$

Finding a violated spanning tree

- For a given $\boldsymbol{\mu}$ find $\max_T I^T(\boldsymbol{\mu}, \boldsymbol{z})$
- If it's positive, add the maximizing tree
- How can we maximize over all trees? Note that:

$$I^T(\boldsymbol{\mu}, \boldsymbol{z}) = \sum_{ij \in T} \Big[\underbrace{\mu_{ij}(z_i, z_j) - \mu_i(z_i) - \mu_j(z_j)}_{w_{ij}} \Big] + \underbrace{\sum_i \mu_i(z_i)}_{\text{Fixed}}$$

- Decomposes into edge scores. Maximizing tree can be found using a maximum-weight-spanning-tree algorithm (e.g., Wainwright 02)

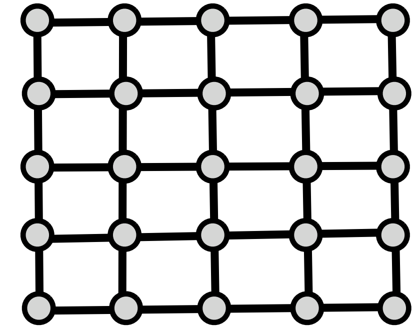
Experiments

- Alternative algorithms for approximate 2nd best:
  - Using approximate marginals from max-product (BMMF; Yanover and Weiss 04)
  - Lawler/Nillson (72,80) - Partition assignments $\boldsymbol{x} \neq \boldsymbol{x}^{(1)}$:

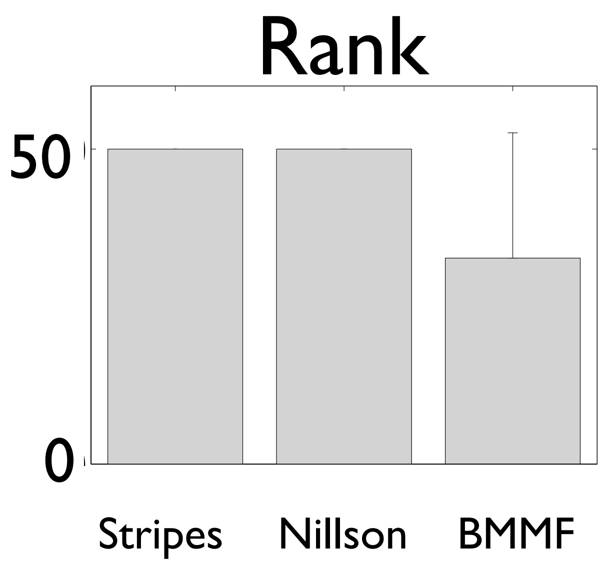
$$
\begin{array}{l|l|l|l|l|}
x_1 \neq x_1^{(1)} & x_2 = * & x_3 = * & \ldots & x_n = * \\
x_1 = x_1^{(1)} & x_2 \neq x_2^{(1)} & x_3 = * & \ldots & x_n = * \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_1 = x_1^{(1)} & x_2 = x_2^{(1)} & x_3 = x_1^{(3)} & \ldots & x_n \neq x_1^{(n)}
\end{array}
$$

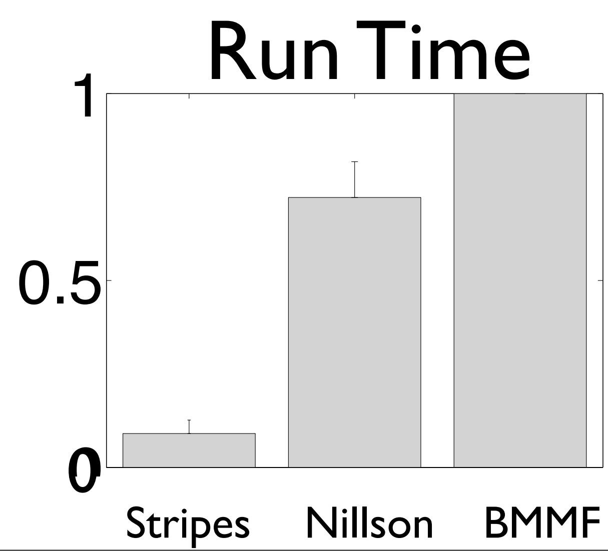
  - Maximize over each part approximately. Cost O(n)
- Our algorithm: STRIPES

Attractive Grids



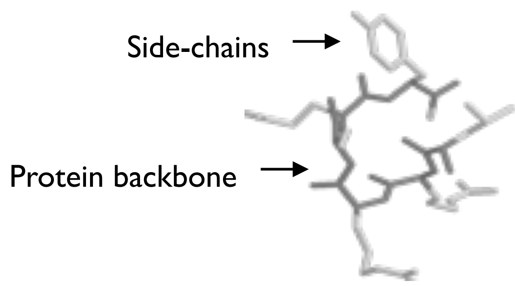
- Ising models with ferromagnetic interaction
- The local-polytope guaranteed to yield exact first best (but not equal to the marginal polytope)
- Goal: Find 50 best. Stripes and Nillson find all of them exactly. Up to 19 spanning trees added



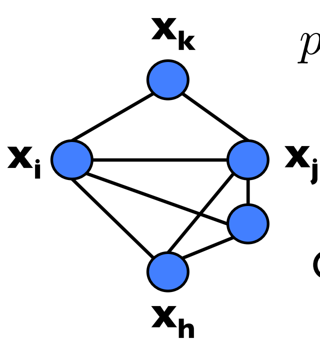


Protein Side Chain Prediction

- Given protein's 3D shape (backbone), choose most probable side chain configuration



(MRFs from Yanover, Meltzer, Weiss '06)

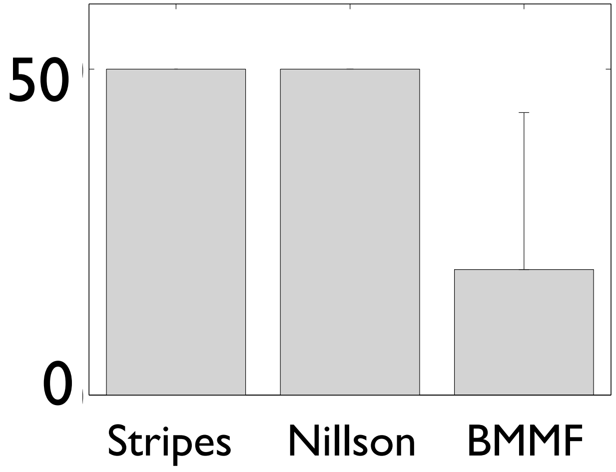


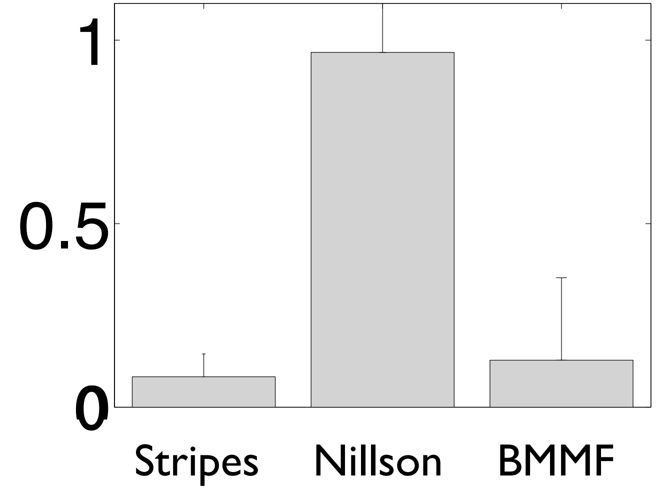
$$p(\boldsymbol{x}) \propto e^{\sum_{ij \in E} \theta_{ij}(x_i, x_j)}$$

- Can be cast as a MAP problem
- Important to obtain multiple possible solutions

Protein Side Chain Prediction

- Stripes found the exact solutions for all problems studied
- In some cases, we used a tighter approximation of the marginal polytope (Sontag et al, UAI 08)





Open Questions

Open Questions

- When are spanning trees enough?

Open Questions

- When are spanning trees enough?
- What is the polytope structure for k-best?

Open Questions

- When are spanning trees enough?
- What is the polytope structure for k-best?
- Finding k-best “different” solutions

Open Questions

- When are spanning trees enough?
- What is the polytope structure for k-best?
- Finding k-best "different" solutions
- Scalable algorithms

Open Questions

- When are spanning trees enough?
- What is the polytope structure for k-best?
- Finding k-best "different" solutions
- Scalable algorithms
- If a given problem is solved with a marginal polytope relaxation, what can we say about the second best?

Open Questions

- When are spanning trees enough?
- What is the polytope structure for k-best?
- Finding k-best "different" solutions
- Scalable algorithms
- If a given problem is solved with a marginal polytope relaxation, what can we say about the second best?

Summary

- The 2nd best can be posed as a linear program
- For trees differs from 1st best by one constraint only
- For non-trees, approximation can be devised by adding inequalities for all spanning trees
- Empirically effective